

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE/DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.



**PCT**

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification <sup>6</sup> :  <b>C12N 15/82, 15/54, A01H 5/00, 5/10</b></p>	<p><b>A1</b></p>	<p>(11) International Publication Number: <b>WO 97/16559</b>  (43) International Publication Date: <b>9 May 1997 (09.05.97)</b></p>
<p>(21) International Application Number: <b>PCT/EP96/04747</b>  (22) International Filing Date: <b>31 October 1996 (31.10.96)</b>  (30) Priority Data:  95402425.3 31 October 1995 (31.10.95) EP  (34) Countries for which the regional or international application was filed: <b>GB et al.</b>  (71) Applicants (for all designated States except US): <b>PLANT GENETIC SYSTEMS, N.V. [BE/BE]; Jozef Plateaustraat 22, B-9000 Gent (BE). NATIONAL RESEARCH COUNCIL OF CANADA [CA/CA]; Montreal Road, Ottawa, Ontario K1A 0R6 (CA).</b>  (72) Inventors; and  (75) Inventors/Applicants (for US only): <b>VAN AUDENHOVE, Katrien [BE/BE]; Duiventorenstraat 20, B-8310 Assebroek (BE). MARILLIA, Elizabeth [FR/CA]; 1205 Spadina Crescent East, Saskatoon, Saskatchewan S7K 3J1 (CA). PEFER-OEN, Mamix [BE/BE]; Berkenlaan 7, B-9850 Nevele (BE). GROOTWASSINK, Jan, Willem, Derk [CA/CA]; 409 Bate Crescent, Saskatoon, Saskatchewan S7H 3A6 (CA). REED, Darwin, Wilfred [CA/CA]; 14 Diggle Place, Saskatoon, Saskatchewan S7M 4J9 (CA). HEMMINGSEN, Sean, Mathias [CA/CA]; 210 Lake Crescent, Saskatoon, Saskatchewan</b></p>		<p><b>S7H 3A2 (CA). KOLENOVSKY, Allan, Dale [CA/CA]; 1202-425 3rd Avenue North, Saskatoon, Saskatchewan S7K 5C4 (CA). UNDERHILL, Edward, Wesley [CA/CA]; 102-1000 Park Boulevard, Victoria, British Columbia V8V 2T4 (CA). MACPHERSON, James, Malcolm [CA/CA]; 18 Madison Crescent, Saskatoon, Saskatchewan S7L 3V8 (CA).</b>  (74) Agents: <b>GUTMANN, Ernest et al.; Ernest Gutmann-Yves Plasseraud S.A., 3, rue Chauveau-Lagarde, F-75008 Paris (FR).</b>  (81) Designated States: <b>AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</b>  Published  <i>With international search report.  Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>
<p>(54) Title: <b>PLANTS WITH REDUCED GLUCOSINOLATE CONTENT</b></p>		
<p>(57) Abstract  <p>The present invention relates to a method to reduce glucosinolate production and/or accumulation in plants wherein a reduced glucosinolate content is desired, particularly in the seeds of these plants. In accordance with the present invention, chimeric gene constructs inhibiting an enzyme responsible for glucosinolate production, the enzyme UDP-glucose:thiohydroximate S-glucosyltransferase (further referred to as "S-GT") are provided. Further in accordance with the present invention, plants are provided having substantially lower glucosinolate levels in their tissues, particularly in their seeds, which permit a significant expansion of the germplasm basis for the breeding of new oilseed rape varieties having lower glucosinolate levels, particularly in their seeds.</p></p>		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

## PLANTS WITH REDUCED GLUCOSINOLATE CONTENT

### **FIELD OF THE INVENTION**

The present invention relates to a method to reduce glucosinolate production and/or accumulation in plants wherein a reduced glucosinolate content is desired, particularly in the seeds of these plants. In accordance with the present invention, chimeric gene constructs inhibiting an enzyme responsible for glucosinolate production, the enzyme UDP-glucose:thiohydroximate S-glucosyltransferase (further referred to as "S-GT") are provided. Further in accordance with the present invention, plants are provided having substantially lower glucosinolate levels in their tissues, particularly in their seeds, which permits a significant expansion of the germplasm basis for the breeding of new oilseed rape varieties having lower glucosinolate levels, particularly in their seeds.

### **BACKGROUND**

Glucosinolates are low molecular weight sulphur-containing glucosides that are produced and stored in almost all tissues of members of the Capparales, the most important member being the group of Crucifer plants (Haughn et al., 1991, Plant Physiol. 97, 217-226). They are composed of two parts, a glycone moiety and a variable a glycone side chain derived from  $\alpha$ -amino acids. When intact, these secondary metabolites are totally passive and innocuous, and do not have any known physiological function.

During the mechanical breaking of the plant tissues, e.g., during consumption, the glucosinolates come into contact with the endogenous enzyme, myrosinase. The resulting breakdown products include isothiocyanates, nitriles, thiocyanates, isocyanates, thiones and alcohols. Intake of large amounts of glucosinolates and their breakdown products causes acute goiter and chronic

diseases in experimental and food-producing animals (Bell, 1993, Can. J. Anim. Sci. 73, 679-697).

From an agronomical point of view, the most important part of the oilseed rape crop are the seeds, which are used to extract the oil. The seed cake is a protein-rich animal feed, which remains after extraction of the oil from the seeds. The presence of glucosinolates in the seed cake is undesired in view of the known toxicity of glucosinolate-breakdown products on animals and humans. Certainly for oilseed crops such as mustard and oilseed rape, the presence of these glucosinolate-breakdown products is problematic. Therefore, breeding programs have been set up to decrease the amount of glucosinolates in these crops, particularly in oilseed rape, to an acceptable level. Unfortunately, glucosinolate production in *Brassica napus* appears to be a multigenic trait.

In Canada, the term "canola" describes oilseed rape with limited levels of glucosinolates and erucic acid in the harvested seeds. Conventional plant breeding efforts have in exceptional cases already resulted in some "zero" glucosinolate oilseed rape varieties.

The cDNA sequence of a *B. napus* myrosinase, an enzyme breaking down glucosinolates, has been described (Falk et al., 1992, Plant Science 83, 181-186). It has been suggested to target the myrosinase to the storage sites of the glucosinolates in the plant cells so as to prematurely destroy them (GrootWassink, 1994a, PBI Bulletin (National Research Council of Canada)).

Furthermore, several attempts have been made to isolate a pure S-GT enzyme from several Crucifers, but these were not successful, since the S-GT enzyme occurs at very low levels, most of the activity is lost upon purification, the enzyme is instable and the enzyme tends to associate with other materials released from the cells (Reed et al. (Arch. Biochem. Biophys. 305, 526-532, 1993); Guo et al. (Phytochemistry 36, 1133-1138, 1994)). Although GrootWassink et al. (1994b, Plant Physiol. 105, 425-433) describe the

immunopurification of S-GT from the florets of Brassica oleraceae spp. botrytis (cauliflower), this method resulted in the isolation of several S-GT isoenzymes with different pI values. Thus, the purification to homogeneity of a sole S-GT enzyme has not been accomplished to date.

It has also been suggested to put the gene for the S-glucosyltransferase enzyme back into canola, but in reversed orientation so that the antisense DNA strand is transcribed (Poulton and Moller, 1993, in Methods in Plant Biochemistry, vol. 9. Academic Press, London, pp. 209-237; GrootWassink, 1994a, supra). However, this suggestion was never accomplished since the isolation or cloning of a DNA sequence encoding an S-GT enzyme has never been reported.

Chavadej et al. (1994, PNAS 91, 2166-2170) expressed a tryptophan decarboxylase in transgenic oilseed rape plants to obtain a significant decrease in the levels of indole glucosinolates. However, the levels of other glucosinolates remained unaltered.

Accordingly, it is an object of the present invention to overcome the problems known in the art by providing a pure S-GT enzyme, the DNA sequence encoding it and a method to reduce glucosinolate production and/or accumulation in plants. These and other objects are achieved by the present invention as evidenced by the summary of the invention, description of the preferred embodiments and the claims.

### **SUMMARY OF THE INVENTION**

Accordingly, it is an object of the present invention to provide a method for reducing the S-GT activity in plant cells, by reducing the expression of the s-gt isoforms genes.

Yet another object of the present invention is to provide plant-expressible chimeric genes and transformation vectors comprising an s-gt inhibitory chimeric gene such as a gene encoding an antisense s-gt RNA. In a preferred embodiment, the present invention provides a plant expressible chimeric gene comprising a gene encoding an antisense RNA complementary to all or part of the RNA encoded by the plant s-gt gene, the sequence of which is comprised in E. coli designated pGT6Sal deposited at the ATCC (American Type Culture Collection, Rockville, Maryland) on October 31, 1996, or of the plant s-gt gene, the cDNA of which is comprised in clone pGL9, which is deposited at the BCCM-LMBP under accesssion number 3344.

Yet another preferred embodiment of the present invention provides a chimeric gene encoding an antisense RNA complementary to all or part (at least 100 nucleotides) of the RNA, a cDNA of which comprises the DNA sequence of SEQ ID No. 28, or an antisense RNA complementary to the RNA encoded by a variant of the s-gt gene, coding for a protein with substantially the same S-GT activity, such as any of the DNA sequences represented in Figure 2.

In yet another aspect, the present invention provides a plant, transformed to contain a chimeric gene comprising:

- a) a plant-expressible promoter,
- b) a transcribed region operably linked to said promoter, comprising a DNA sequence encoding an RNA or protein, wherein said RNA or protein interfere with the normal expression of the UDP-glucose:thiohydroximate S-glucosyltransferase gene (s-gt gene) in cells of said plant, and
- c) a 3' transcription termination and polyadenylation region active in said plant, as well as seeds, and seed cakes obtained from said seeds comprising said transcribed region b).

Yet another aspect of the present invention provides a DNA comprising a region encoding a protein with UDP-glucose:thiohydroximate S-glucosyltransferase activity, selected from the following groups:

- 1) a DNA encoding an mRNA, the cDNA of which is contained in plasmid pGL9, deposited in E. coli WK6 at the BCCM-LMBP under accession number 3344;
- 2) a DNA encoding an mRNA, the cDNA of which has the sequence of SEQ ID No.28;
- 3) a DNA having substantial sequence homology or similarity to SEQ ID No. 28; and
- 4) a DNA with the sequence of SEQ ID No. 34.

Yet another aspect of the present invention provides a DNA sequence encoding an antisense RNA selected from the following groups:

- 1) an antisense RNA which is complementary, preferably at least 90 % complementary, more preferably at least 95 % complementary, to a region of at least 500 nucleotides, of an mRNA, the cDNA of which is contained in plasmid pGL9, deposited in E. coli WK6 at the BCCM-LMBP under accession number 3344;
- 2) an antisense RNA, which is at least 90 % complementary, more preferably at least 95 % complementary, to a region of at least 100 nucleotides, preferably a region of at least 500 nucleotides, of an mRNA, the cDNA of which comprises the coding region of SEQ ID No. 28;
- 3) an antisense RNA encoded by the s-gt inhibitory gene contained in plasmid TKV8a included in E. coli MC1061, deposited at the BCCM-LMBP under accession number LMBP 3343, or an RNA having substantial sequence similarity thereto; and
- 4) an antisense RNA encoded by the s-gt inhibitory gene contained in E. coli, deposited at the ATCC on October 31, 1996, or an RNA having substantial sequence homology thereto.

Yet another aspect of the present invention provides a process for obtaining a Brassica napus plant having a significantly reduced expression of an s-gt gene, comprising the following steps:

- a) transforming a plant cell with a s-gt inhibitory chimeric gene; and



- b) regenerating a plant from said transformed cell.

Also contemplated by the present invention are hybrid plants having a glucosinolate content of less than 30  $\mu$ moles per gram dry defatted seed glucosinolates in their seeds.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

Fig.1 is a schematic representation of the different cDNA clones obtained from B. napus S-GT mRNA and their delineation on the full pGL9 cDNA obtained. The thick lines represent the coding region, the thinner lines represent the 5' and 3' untranslated sequences of the cDNA. The primers used to obtain the cDNAs are indicated, with their corresponding direction ("ANCH" refers to the 5' or 3' Anchor primers (Clontech) described below, locations are approximately and the arrows representing the primers are not drawn to scale).

Fig. 2 is the ORF of the pGL9 clone of SEQ ID No. 28 indicating all nucleotide differences found in the coding regions of the different s-gt cDNAs isolated in accordance with this invention. The altered nucleotide in the coding regions of the other cDNAs is indicated above the DNA sequence of the pGL9 clone, the corresponding nucleotide in the DNA sequence of pGL9 is in lowercase letters. The numbers between brackets are as follows: (6) refers to the pGL6-14 clone, (3) to the pGL3-22, (4) to pGL4-2, (7) to pGL2-7, (25) to pGL2-25. The underlined parts were found to be identical between pGL18 clone 1 and the pGL9 clone. No amino acid differences are indicated in the Figure. The 5' and 3' ends of the open reading frames contained in the different cDNA clones are also shown above the pGL9 sequence by the marks < and >, respectively (the consecutive numberings refer to amino acid positions).

### **DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS OF THE INVENTION**

This invention provides a method to reduce glucosinolates production and/or accumulation in plants. This method was achieved by isolating DNA sequences encoding a Brassica napus S-GT enzyme form. Starting from the partial amino acid sequences of peptide fragments obtained from the purified B. oleraceae S-GT, degenerate DNA primers were designed and DNA fragments encoding all or part of an S-GT enzyme were isolated by PCR-RACE. Also, based on these partial amino acid sequences, the B. napus genomic clone encoding an S-GT enzyme was isolated.

The following definitions are provided to further clarify the terminology used throughout the specification, it being understood that these definitions are provided for the person skilled in the art to solely use as a basis for interpreting the preferred embodiments, the examples and the claims.

As used herein, "substantial sequence similarity", refers to DNA sequences encoding similar RNAs and/or proteins with some differences in their RNA or amino acid sequence, e.g., nucleotide or amino acid deletions, additions, or replacements, with the proviso that these similar RNAs and/or proteins still retain significantly the same function or activity. Preferably, DNA sequences with "substantial sequence similarity" encode proteins having the same tertiary structure in those domains or regions determining the protein activity. Also encompassed in the definition of "substantial sequence similarity", when referring to DNA sequences, are different DNA sequences encoding the same proteins. Indeed, because of the degeneracy of the genetic code, many different DNA sequences can encode one protein. Also, during cloning work some nucleotides can be changed to create suitable restriction sites throughout a DNA sequence, or introns can be inserted, while retaining substantial sequence similarity. These nucleotide changes made during cloning are also encompassed by the term "substantial sequence similarity".

Furthermore, natural variants having substantial sequence similarity to a DNA s quence differing in some nucleotides or synthetic variants having

substantial sequence similarity which can be made by recombinant DNA techniques are encompassed by the definition of "substantial sequence similarity". It is to be understood that the terms "homology" and "similarity" can be used interchangeably in the context of the present invention.

In a preferred embodiment of this invention, DNA sequences have substantial sequence similarity if they have more than 85 %, preferably more than 90 %, more preferably more than 95 %, sequence similarity.

Sequence similarity between two nucleotide sequences is conveniently measured by the Wilbur and Lipmann algorithm with the IntelliGenetics™. (Intelligenetics Inc.) sequence analysis package (Wilbur and Lipmann, 1983, Proc. Natl. Acad. Sci. USA 80, 726) using a window-size of 20 nucleotides, a word length of 4 nucleotides and a gap penalty of 4.

It is known that some amino acids in a protein can be replaced by others, provided the tertiary structure is not significantly altered. Therefore, for proteins, "substantial sequence similarity" refers to proteins differing in some amino acids, e.g., by amino acid deletions, additions or replacements, while retaining the same overall function as determined by the proteins tertiary structure. Typically, amino acid differences between functionally active protein forms with substantial sequence similarity are below 5 %, preferably below 3 %, of the total number of amino acids. Sequence similarity between two protein sequences can be conveniently measured by the Wilbur and Lipmann algorithm with the IntelliGenetics™ (Intelligenetics Inc.) sequence analysis package (Wilbur and Lipman, 1983, supra) using a window size of 20 amino acids, a word length of 2 amino acids, and a gap penalty of 4. Proteins wherein amino acids are replaced by conservative amino acids with similar physicochemical characteristics are also included in the definition of proteins with substantial sequence similarity.

A "chimeric gene", as used herein, is a gene wherein at least one regulatory region is heterologous to and therefore not normally associated with

the coding region or the transcribed region in nature; e.g., an s-gt coding region under the control of a bacterial promoter or a plant-expressible promoter of another plant gene.

As used herein, "a plant-expressible chimeric gene" is a gene expressible in cells of a plant, comprising at least one regulatory region, e.g., a plant-expressible promoter or a 3' transcription termination and polyadenylation sequence active in plant cells, which is not normally associated with the transcribed region or the coding region in plant cells in nature. A plant-expressible chimeric gene in accordance with this invention typically encodes an RNA which is translated into a protein, or an RNA which is functional as such, e.g., an antisense RNA or a ribozyme. A DNA encoding an (antisense) RNA, complementary to a (sense) RNA produced in a plant cell, operably linked to the promoter and 3' transcription termination and polyadenylation region normally regulating transcription of the (sense) RNA, is also comprised under the definition of chimeric gene in accordance with this invention.

As used herein, "plant-expressible promoter" is a promoter active in plant cells, including but not limited to promoters of plant origin and of bacterial or viral origin that are functional in plant cells (e.g., CaMV 35S promoter, Agrobacterium T-DNA promoters and the like). Examples of viral promoters include, but are not limited to, viral promoters that can transcribe RNA in a plant cell when an appropriate polymerase is also expressed in the same plant cell, such as those described by Lasstner et al. (1991, Plant Mol. Biol. 17, 229-234).

As used herein, "promoter" is a nucleotide sequence recognized (directly or indirectly) and bound by DNA-dependent RNA polymerase during initiation of transcription.

As used herein, "promoter region" is a DNA sequence typically located 5' of a coding region, and including the promoter and a 5' untranslated leader sequence.

As used herein, "transcribed region" is that region of a DNA transcribed into an RNA, typically comprising 5' leader sequences, a coding region (which, by definition includes a region encoding a protein or a region encoding an antisense RNA or a ribozyme) and 3' untranslated trailer sequences.

As used herein, "antisense RNA" is an RNA molecule which, by binding to (hybridizing to) a complementary sequence in another nucleic acid molecule (RNA or DNA), inhibits the function and/or completion of synthesis of the complementary molecule.

Antisense RNA is typically produced from a chimeric gene by operably linking all or part of a DNA sequence encoding an RNA sequence to a promoter in the orientation opposite to the orientation in which the DNA encoding the RNA (or its part) is operably linked to its promoter in an endogenous plant gene, so that an RNA is formed which is complementary to all or part of the RNA normally produced from the endogenous gene in a plant cell. Antisense RNA can comprise a sequence complementary to all or part of the 5' and 3' untranslated regions of an RNA to be inactivated, or even a sequence complementary to introns or parts of introns of a pre-mRNA, such that the antisense is rendered more specific. The antisense RNA is at least 85 %, preferably at least 90 %, most preferably 95 to 100 %, complementary to the RNA to be inhibited.

The term "complementary", as used herein, refers to a sequence of nucleotide bases in one strand of a DNA or RNA molecule that is exactly complementary to that on another strand such that there is no variation between adenine-thymine, adenine-uracil or guanine-cytosine base pairs.

As used herein, "90 % complementary" refers to an RNA sequence wherein 90 % of the nucleotides are complementary to the corresponding nucleotides in another RNA sequence, preferably of the same length, so that only 10 % of the nucleotides will not hybridize to the corresponding nucleotide on the

other RNA or DNA. For example, for two RNAs with the same number of nucleotides, when RNA sequence 1 is 90 % complementary to an RNA sequence 2, the complementary form of an RNA sequence 1 will have 90 % sequence similarity with RNA sequence 2.

"S-GT", or "S-GT protein" as used herein, refers to a member of the UDP-glucose:thiohydroximate S-glucosyltransferase enzyme family (EC 2.4.1.-) of plants, particularly Brassica plants, that comprises the enzyme forms which catalyze the penultimate step in glucosinolate production. A preferred S-GT protein in accordance with this invention is the protein with the amino acid sequence of SEQ ID No. 35.

"s-gt gene" or "s-gt DNA", as used herein, refers to a DNA sequence encoding an S-GT protein. A preferred s-gt DNA in accordance with the present invention includes that DNA in SEQ ID No. 34.

As used herein, "functionally equivalent parts" of a DNA, RNA or protein are portions of a DNA, RNA or protein which have the same function or activity as the full DNA, RNA or protein. For example, a functionally equivalent part of an antisense RNA, is a portion of an antisense RNA, wherein this portion exhibits substantially the same antisense (inhibitory) effect as the full length antisense RNA, although the portion will differ from the full length antisense RNA in certain other characteristics such as molecular weight, its size, its relative nucleotide composition and the like. Also, a protein fragment, which has the same metabolic activity (e.g., glucosyltransferase activity in plant cells) as the entire protein, is a "functionally equivalent part" of that protein in accordance with this invention.

As used herein, "gene silencing" is a significant or complete reduction in detectable gene expression. Gene silencing can be achieved at any level of gene expression, and results in a significant drop of production of RNA or

protein. In a preferred embodiment, gene silencing is achieved by using a DNA encoding an antisense RNA or a functionally effective part thereof.

As used herein, "partial gene silencing" refers to an inhibition of expression of less than 100 %, preferably from about 50 to about 90 %, as measured by RNA or proteins levels, while "complete gene silencing" refers to the situation wherein no detectable gene expression (RNA or protein) is observed.

As used herein, a "ribozyme" is a catalytic RNA molecule capable of specifically cleaving another RNA. A ribozyme typically has a "targetting" sequence which is complementary to another RNA, so that the ribozyme can recognize and cleave this other RNA. The targetting sequence preferably is at least 90 %, more particularly at least 95 %, preferably 100 %, complementary to the RNA which is to be cleaved.

As used herein, the terms "significantly reduced", "significant inhibition", or "significantly lower levels", when referring to s-gt gene expression, refer to a quantitative difference in expression of the native s-gt gene in a plant cell transformed with an s-gt inhibitory chimeric gene when compared to the situation in the wild-type plant, as is evidenced by protein levels measured via quantitative protein assays (e.g., ELISA) in a plant cell. Preferably, significantly reduced or inhibited expression of an s-gt gene in accordance with this invention refers to a reduction in formation of S-GT protein of 50 % to 95 %, particularly at least 75 %, more particularly at least 85 %, preferably 95 %, in a plant cell as is measured by protein quantitative assays in comparison to a cell of the same cell type used as control, or by S-glucosyl transferase activity as measured by GrootWassink et al., 1994b, supra.

"Reduced total glucosinolate", as used herein, refers to a reduction in total glucosinolate level is to below 30  $\mu$ moles per gram oil-free seed matter, preferably below 10, more preferably below 5  $\mu$ moles per gram oil-free seed matter,

particularly to undetectable levels using the method of GrootWassink et al. (1994b, supra).

In accordance with this invention, DNA sequences encoding S-GT proteins have been isolated. All or part of the isolated s-gt DNA, preferably the s-gt coding region, can be used in a variety of ways, including but not limited to the production of S-GT protein in bacteria, the use of the isolated s-gt DNA in chimeric genes, plants transformed by the chimeric genes according to the invention, the production of seeds lacking or having reduced content of glucosinolates and the production of seed cakes obtained by the crushing of the seeds.

To produce S-GT protein free from contaminating plant proteins, all or part of the isolated s-gt DNA and preferably the s-gt coding region can be used. Preferred promoter and 3' transcription termination sequences for the chimeric s-gt gene to be expressed in bacteria are derived from bacterial genes (see, Sambrook et al., Molecular cloning. A laboratory Manual (1989)). The s-gt cDNA which corresponds to the full open reading frame of an s-gt gene (and which is contained in plasmid pGL9 (BCCM-LMBP 3344)), was cloned into an E. coli expression vector, by methods known in the art, to recombinantly produce the protein. The activity of the recombinantly produced S-GT enzyme was confirmed by assaying for glucosyl-transferase-activity using the assay as described by Reed et al. (1993, supra) and GrootWassink et al. (1994b, supra). Similarly, the protein produced from the coding region contained in the genomic clone designated pGT6Sal deposited at the ATCC on October 31, 1996, is confirmed to have glucosyl-transferase activity in the same assay.

Preferred DNA sequences encoding a B. napus S-GT enzyme are shown in SEQ ID No. 28 and in SEQ ID No. 34. Nevertheless, other different isoforms of the gene exist, having some amino acid differences. Some amino acids were found to be different in the B. napus clone when compared to the peptide fragments obtained from the B. oleraceae S-GT form. As is clear from the



Examples and from Figure 2, also in B. napus, evidence for the existence of isozymes differing in some amino acid residues was found. This indicates that a small family of related s-gt genes having substantial sequence similarity exists in Brassica plants.

The following amino acids were found to be different between the S-GT protein of SEQ ID No. 28 and the other isoforms, of which a cDNA clone was isolated. The amino acid which is different at this position in an isoform, and between brackets, the amino acid in SEQ ID No. 28 are indicated below: position 2: Val(Ala); between position 10-11: add Lys between amino acids 10 and 11; position 12: Ser(Asn); position 43: Leu(Val); position 75: Pro(Leu); position 88: Gly(Glu); position 93: His(Asn); position 96: Gln(Glu); position 133: Leu(Ile); position 153: Ala(Val); position 167: Leu(Pro); position 204: Ile(Arg); position 216: Gly(Ser); position 232: Thr(Ala); position 234: Lys(Arg); position 49: Ala(Gly); position 290: Arg(Gly); position 302: Thr(Lys); position 319: Arg(His); position 350: Gly(Val); position 350: Glu(Val); position 402: Asp(Glu); position 419: Lys(Arg). Preferred amino acid differences in such isoforms include the above amino acid differences at the positions 75, 153, 216, 234, and 350.

Therefore, preferred S-GT proteins of the invention include variants of the protein of SEQ ID No. 28 having at least one of the following amino acids in the protein of SEQ ID No. 28 changed into another amino acid indicated in the above list.

Preferably, the Cysteine amino acids are not altered in S-GT enzymes having substantial sequence similarity to the S-GT form encoded by the cDNA contained in clone pGL9, since SH-bonds are expected to be involved in enzymatic activity. Also, preferred s-gt DNA's or coding regions in accordance with this invention are those DNA sequences encoding the S-GT variants, having at least one of the above amino acid substitutions. A functionally equivalent variant or part of an S-GT enzyme of SEQ ID No. 28, in accordance with this invention, is an enzyme having at least 80 %, preferably at least 90 %, of the

activity of this S-GT enzyme in the glucosyltransferase radioassay of GrootWassink et al. (1994b, *supra*), when analyzed under the same test conditions.

Furthermore, given the DNA sequence of the S-GT form isolated, variants can be designed having a different codon usage or containing additional untranslated sequences such as introns but encoding the same protein or a protein having substantial sequence similarity. Also, based on the amino acid sequence of the protein encoded by the cDNA comprised in clone pGL9, variants can be made differing in some amino acids but having substantially the same glucosyltransferase activity. Upon initial comparison of the protein primary structure with other plant glucosyltransferases, particularly the protein parts from amino acids 16-42, 180-185, 346-375, 448-453 in SEQ ID No. 28 or SEQ ID No. 34 should be retained when designing variants of the B. napus S-GT enzyme. Therefore, amino acid substitutions in these regions should be kept to minimal (e.g., no more than 10 %, preferably no more than 5 %) in order to retain most of the S-GT activity.

In a preferred embodiment of this invention, DNA sequences with substantial sequence similarity to the s-gt DNA of the invention include DNA sequences encoding the protein produced by plasmid pGL9 contained in E. coli WK6, deposited in accordance with the Budapest Treaty at the BCCM-LMBP on September 7, 1995 under accession number 3344. Particularly, DNA sequences with substantial sequence similarity include DNA sequences encoding the protein produced by the deposited clone designated pGT6Sal deposited at the ATCC on October 31, 1996.

Variants of the S-GT enzyme can be isolated based on the knowledge of the protein peptidic fragments of the B. oleraceae S-GT enzyme of the invention. A S-GT enzyme variant of the present invention should retain the enzyme's catalytic site and thus is characterized by its clear S-GT activity in the radioassay of Reed et al. (1993, *supra*) and the presence of a continuous stretch of amino

acids with more than 85 % sequence similarity, particularly more than 90 % sequence similarity, preferably more than 95 % sequence similarity, to any one of the peptide fragments 1 to 7 of SEQ ID Nos. 1 to 7. Preferentially variants of the S-GT enzyme of this invention have S-GT activity and comprise any one of the peptides of SEQ ID Nos 1 to 7, preferably the peptide of SEQ ID No. 1, in their primary sequence. Since it was found that none of the peptides of SEQ ID Nos 1 to 7 are found in any known protein, the presence of any of these peptides in a protein sequence having glucosyltransferase activity characterizes the S-GT protein of the invention.

Generally in accordance with the present invention, s-gt inhibitory chimeric genes are provided for plant transformation. As used herein, "s-gt inhibitory chimeric gene" is a chimeric gene comprising all or part of a coding sequence encoding an RNA or protein inhibiting s-gt gene expression. For example, a chimeric gene of the present invention includes all or part of a DNA encoding an antisense RNA, a sense RNA, or a ribozyme, inhibiting s-gt gene expression when produced in a plant cell. An s-gt inhibitory chimeric gene typically also comprises, besides a coding region, a promoter region and a 3' transcription termination and polyadenylation region.

Any plant expressible promoter can be used to express the inhibitory chimeric gene of the present invention.

Preferred plant-expressible promoters of the s-gt inhibitory chimeric gene of this invention include, but are not limited to: the strong (constitutive) 35S promoters (the "35S promoters") of the cauliflower mosaic virus of isolates CM 1841 (Gardner et al., 1981, Nucleic Acids Research 9, 2871-2887), CabbB-S (Franck et al., 1980, Cell 21, 285-294) and CabbB-JI (Hull and Howell, 1987, Virology 86, 482-493); the ubiquitin promoter (EP 0342926); viral promoters and their polymerase as described in Lasstner (1991, supra), and the TR1' promoter and the TR2' promoter which drive the expression of the 1' and 2' genes,

respectively, of the Agrobacterium tumefaciens T-DNA (Velten et al., 1984, EMBO J. 3, 2723-2730).

Alternatively, a promoter can be utilized which is not constitutive but rather is specific for one or more tissues or organs of the plant, such as the pod tissue, whereby the inserted chimeric s-gt inhibitory gene or its functionally effective part is expressed only or mostly in cells of the specific tissue or organ. Since it has been shown that the important glucosinolates stored in the seed are derived from the adjacent pod tissue, particularly from the pod walls (Toroser et al., 1995 Annual Meeting of the American Society of Plant Physiologists, Charlotte, North Carolina, July 29-August 2, USA, in Supplement to Plant Physiology, vol. 108, abstract 716), a plant-expressible promoter can be utilized which is at least expressed in pod tissue, and particularly in the pod wall; preferably highly expressed in pod tissue, more preferably the pod wall. Pod tissue-specific promoters as used herein include the promoters of the genes encoding pod-specific mRNAs, e.g. the promoters of the genes encoding the mRNAs reported by Coupe et al. (1993, Plant Mol. Biol. 23, 1223; 1994, Plant Mol. Biol. 24, 223).

For example, preferential expression in the pod tissue by a pod tissue-specific promoter, preferably in the pod wall (e.g., the carpels constituting the pod wall) by a pod-wall-specific promoter, is an alternative method to lower seed glucosinolate levels without interfering, to a large extent, with the glucosinolate content of the leaves and other tissues, which may be more desirable in certain circumstances.

"Pod tissue", as used herein, refers to the cells constituting a pod, including structures such as the carpels forming the pod wall, the (false) septum, and the replum but excluding the seeds.

In yet another embodiment of the present invention, the promoter for the s-gt inhibitory chimeric gene of the invention can also be the promoter of the endogenous s-gt gene, the mRNA of which corresponds to the cDNA contained

in plasmid pGL9, or can be a leaf-specific, stem-specific and in some cases even an embryo- or seed-specific promoter. In one embodiment of this invention, the promoter region is characterized by part of the DNA sequence of SEQ ID No. 34 from position 1 to position 212, i.e., the upstream sequences of the s-gt coding region.

In the present invention, for example, the upstream region between 3 to 5 Kb and 0.5 Kb upstream of the ATG translation initiation codon are useful as a promoter region, particularly the upstream region between 2.5 kb and 0.5 kb upstream of the ATG translation initiation codon of the DNA of SEQ ID No. 34.

Furthermore, a comparison of the s-gt DNA sequences shows that differences between the s-gt DNA forms are more different in the 5' untranslated sequences, including the promoter region, than in the coding region. Thus, in the present invention the different promoter regions of the s-gt forms are isolated from a genomic library so as to use them in the different approaches of expression of an s-gt inhibitory gene in accordance with this invention.

In accordance with this invention, the s-gt inhibitory chimeric gene, or a functionally effective part thereof, is inserted in the plant genome so that the inserted coding region is upstream (i.e., 5') of suitable 3' transcription termination and polyadenylation signals (i.e., transcript termination and polyadenylation signals). Preferred polyadenylation and transcript formation signals include the CaMV 35S polyadenylation and transcript formation signals (Mogen et al., 1990, The Plant Cell 2, 1261-1272), and those of the octopine synthase gene (Gielen et al., 1984, EMBO J 3, 835-845) and the T-DNA gene 7 (Velten and Schell, 1985, Nucl. Acids Res. 13, 6981-6998), which act as 3'-untranslated DNA sequences in transformed plant cells. Alternatively, the 3' transcription termination and polyadenylation signals can be obtained from a pod tissue or pod wall specific gene or from a plant s-gt gene.

More specifically, the coding region of the s-gt inhibitory chimeric gene of the invention comprises a DNA encoding an RNA or protein interfering with the normal expression of the s-gt gene in the plant cell. A preferred coding region in accordance with this invention comprises a DNA sequence encoding an s-gt antisense RNA, or functionally effective variants thereof. In accordance with this invention, "s-gt antisense RNA" is an RNA complementary to at least part, preferably a functionally effective part, of the RNA, particularly the mRNA, encoded by an endogenous plant s-gt gene. For example, an RNA complementary to a region of at least 100 nucleotides, preferably a region of at least 500 nucleotides, of the (m)RNA encoded by the s-gt gene, the cDNA of which is contained in pGL9 (deposited at the BCCM-LMBP under number LMBP 3344) can be utilized. An RNA complementary to a region of at least 100 nucleotides, preferably a region of at least 500 nucleotides, of the (m)RNA encoded by the s-gt gene of SEQ ID No. 34 is also encompassed by this invention. Similarly, as described above for the s-gt DNA sequence, variants of the s-gt antisense RNA can be made, provided they have sufficient complementarity, preferably at least 90 % complementarity, particularly at least 95 % complementarity, to the s-gt RNA formed in the plant cell.

Surprisingly, it was found that there are several isoform DNAs of the invention which have substantial sequence similarity, so one s-gt antisense RNA will be able to inhibit expression of all isoform genes. The antisense technology is based on blocking the information flow (by transcription or translation) from DNA to RNA and/or protein by the introduction of an RNA strand ("antisense RNA") complementary to the sequence of an endogenous target ("sense") RNA (Murray & Crockett, 1992, in Antisense RNA and DNA, pp. 1-50, Murray, JAH (ed.), Wiley-Liss, New York)). The outcome of this is a partial to complete silencing of endogenous gene expression (see, e.g., EP 0 467 349; EP 0 223 399; EP 0 240 208).

It has been shown that even the most abundant protein present in plants can be reduced effectively using antisense techniques (Rodermel et al., 1988,

Cell 55, 673; Jiang et al., 1994, Plant Mol. Biol. 25, 569-576). The antisense RNA can be complementary to only part of the (m)RNA transcribed from the gene that is to be inactivated. Preferably, such a part is at least about 100 basepairs long, more preferably at least about 500 basepairs long, typically about 500-1000 basepairs long. Also an antisense RNA which is complementary to the full (m)RNA to be inactivated, or to the 5' and 3' untranslated regions, can be used.

Also, in view of the down-regulation desired, the antisense RNA can be complementary to certain stretches of conserved sequences or to the most divergent sequences between the different s-gt RNA forms. Hence, the expression of a specific s-gt isoform, or of several isoforms sharing a conserved region, can also be inhibited. For isoform-specific inactivation of the s-gt gene, the 5' leader and 3' trailer sequences will be more useful target sequences, since it has been shown in the present invention that the S-GT isoform DNAs differ most in these regions, while differences in the coding region are limited. When different antisense RNAs can be used, it is preferred to use the antisense RNA which is most stable in the plant cell, particularly in the nucleus.

A preferred antisense construct of this invention is a construct comprising a DNA sequence encoding an RNA which is complementary, preferably 90 % to 100 % complementary, to at least 100 nucleotides of the RNA encoded by the s-gt gene, the mRNA of which corresponds to the cDNA sequence contained in plasmid pGL9 (deposited in host organism E. coli WK6 under accession number BCCM-LMBP 3344), or to the RNA encoded by the s-gt gene corresponding to the DNA sequence of SEQ ID No. 34. It is more preferably to utilize a DNA sequence encoding an RNA which is complementary, preferably 90 % to 100 % complementary, to at least 100 nucleotides up to the entire mRNA sequence encoded by the s-gt gene, wherein a cDNA prepared from this mRNA has the sequence shown in SEQ. ID No. 28, or has substantial sequence similarity to the sequence of SEQ ID No. 28, such as a DNA sequence encoding an RNA which

is complementary to the RNA, preferably the mRNA, encoded by the DNA of SEQ ID No. 34.

Preferred DNA sequences encoding an antisense s-gt RNA, in accordance with this invention include DNA sequences encoding those RNAs hybridizing under stringent conditions with the (m)RNA (or parts thereof) encoded by the s-gt gene in a plant cell. This antisense RNA is at least 80 %, more particularly at least 90 %, more preferably at least 95 % complementary to the s-gt (m)RNA encoded by the DNA sequence of SEQ ID No. 34, particularly to the s-gt mRNA encoded by the s-gt gene, the cDNA of which is contained in plasmid pGL9 (BCCM-LMBP 3344). Indeed, it has been shown that for antisense inhibition of gene expression, 100 % complementarity is not required, however the antisense RNA has to be sufficiently complementary to the sense RNA so that they can readily hybridize under stringent conditions. This way, an entire gene family having high sequence similarity can be inhibited by expressing one antisense RNA which is 100 % complementary to one of the members of the gene family (see, e.g., Rodermeil et al., 1988, supra).

In a further embodiment of this invention, the coding region of the chimeric s-gt inhibitory gene encodes a sense RNA which is identical, preferably has more than 80 % sequence similarity, more preferably has more than 90% sequence similarity, most preferably has at least 95 % sequence similarity, to at least part of the (m)RNA transcribed from the gene, the cDNA of which corresponds to the DNA sequence contained in plasmid pGL9, preferably the DNA with the sequence of SEQ ID No. 34, or functionally effective parts thereof. Similarly as for the antisense approach, a 5' or 3' part of the coding region, the full coding region as well as the 5' or 3' untranslated regions or the intron sequences can be used for gene silencing. Although antisense suppression is the preferred embodiment of this invention, the sense suppression (or co-suppression) mechanism (Flavell et al., 1994, Proc. Natl. Acad. Sci. USA 91, 3490-3496) can also yield plants having significant inhibition of s-gt gene expression in selected transformants.



Alternatively, the coding region of the chimeric s-gt inhibitory gene comprises a DNA sequence encoding a catalytic RNA molecule called a ribozyme (PCT publication WO 89/05852; Nature 334, 585-591, 1988), comprising a targetting region which is complementary, preferably 90 % to 100 % complementary, to part of the RNA formed by an s-gt gene, the mRNA of which corresponds to the cDNA contained in plasmid pGL9, or functionally effective variants thereof. The use of ribozymes allows the targetting of conserved or divergent RNA sequences, such that the expression of a group of S-GT isoforms (sharing a conserved region) or a particular S-GT isoform can be inhibited. Similarly, the EGS technology (e.g., PCT patent publication WO 93/22434) can be used to specifically inactivate s-gt gene expression.

The chimeric s-gt inhibitory gene can also encode an RNA encoding a protein inhibiting S-GT activity, such as an antibody fragment specifically binding to an S-GT, particularly the S-GT encoded by the DNA sequence comprised in clone pGL9, contained in E. coli WK6 (BCCM-LMBP 3344), more preferably the S-GT with the amino acid sequence of SEQ ID No. 28 or SEQ ID No. 34. Such antibodies or the Fab fragments thereof with high affinity to the S-GT protein isolated from clone pGL9 can be expressed in plants (Tavladoraki et al., 1993, Nature 366, 469-472), so that the endogenous S-GT protein is rendered inactive.

Since the S-GT enzyme was found to be non-specific for the side chain of the glucosinolate produced (Jain et al., 1988, J. Plant Physiol. 136, 356-361) the concentrations of all glucosinolates, and not just one type, normally produced in the plant, will be lowered when expressing a chimeric s-gt inhibitory gene, preferably a DNA sequence encoding an s-gt antisense RNA, in plants.

In order to transfer all or a functionally effective part of a chimeric s-gt inhibitory gene to a plant cell genome, suitable restriction sites can be introduced, flanking the chimeric gene or its part. This can be done by site-directed mutagenesis, using well-known procedures (e.g., Stanssens et al., 1989,

Nucl. Acids Res. 12, 4441-4454; White et al., 1989, Trends in Genet. 5, 185-189).

In another embodiment of the invention, plant cells are transformed with the above chimeric genes, and plants are regenerated from such transformed cells. A disarmed Ti plasmid, containing the chimeric s-gt inhibitory gene, in Agrobacterium tumefaciens can be used to transform the plant cell, and thereafter, a transformed plant can be regenerated from the transformed plant cell using the procedures described, for example, in EP 0116718, EP 0270822, PCT publication WO 84/02913 and EP 0242246 (which are also incorporated herein by reference), and in Gould et al. (1991, Plant Physiol. 95, 426-434), or the method described in PCT publication WO 94/00977. Of course, other types of vectors can be used to transform the plant cell, using procedures such as direct gene transfer (as described, for example in EP 0233247), pollen mediated transformation (as described, for example in EP 0270356, PCT publication WO 85/01856, and US Patent 4,684,611), plant RNA virus-mediated transformation (as described, for example in EP 0067553 and US Patent 4,407,956), and liposome-mediated transformation (as described, for example in US Patent 4,536,475).

A resulting transformed plant, such as a transformed oilseed rape plant, can be used in a conventional plant breeding scheme to produce more transformed plants with the same characteristics or to introduce the chimeric s-gt inhibitory gene in other varieties of the same or related plant species or into commercial hybrid plants. Seeds, which are obtained from the transformed plants, contain the chimeric s-gt inhibitory gene as a stable genomic insert.

The preferred plants to be transformed in accordance with this invention include but are not limited to Crucifer plants, particularly Brassica plants, preferably oilseed rape Brassica plants, more preferably Brassica napus plants. The plants of the invention are characterized by their significantly lower levels of expression of the endogenous s-gt gene, preferably these plants are

characterized by the lower glucosinolat content in their tissues, particularly in their seeds.

Hybrid plants, in accordance with this invention, can be made by crossing a transformed plant of the invention, homozygous for the s-gt inhibitory gene, with a male sterile plant, wherein the male sterile plant preferably is obtained by following the method described in EP 344029 and by Mariani et al. (1990, Nature 347, 737-741; 1992, Nature 357, 384-387).

In yet another preferred embodiment of the invention, the obtained hybrid plants containing the chimeric s-gt inhibitory gene correspond to the "canola" definition set out by the Canola Council of Canada. For oilseed rape to be designated "canola", oil obtained after crushing the seed has to contain less than 2 % of total fatty acids in the oil as erucic acid and seed meal derived from said crushed seeds has to contain less than 30  $\mu$ moles of alkenyl glucosinolates per gram of dry matter of the (oil-free) seed meal. Therefore, the eventual hybrid plants obtained in accordance with this invention preferably have a content of alkenyl glucosinolates, preferably total glucosinolates, of less than 30  $\mu$ moles, preferably less than 15  $\mu$ moles, particularly less than 5  $\mu$ moles per gram defatted dry matter of the seed. In the most preferred embodiment of this invention, the plants contain less than 0.5  $\mu$ moles total glucosinolates, preferably alkenyl glucosinolates, per gram dry matter of defatted seed, when the plants are hemizygous for the introduced chimeric s-gt inhibitory gene.

In yet another embodiment of this invention, those tissues or organs of the plant other than pods or seeds can still contain their "wild-type" glucosinolate content. However, the total glucosinolate content, preferably the alkenyl glucosinolate content, in the entire plant is reduced to the above cited levels, preferably to undetectable levels in the most preferred embodiment.

Tentatively, promoters specific for leaves or other plant tissues or organs can be used to lower the glucosinolate content in these tissues or organs while maintaining the "wild-type" glucosinolate levels in other parts of the plant.

In a preferred embodiment of this invention, the glucosinolate content in the plants, preferably in the seeds, is significantly decreased in hybrids, obtained from cross-fertilization of two parent plants, at least one of which has been transformed to inhibit s-gt gene expression in accordance with this invention. Further in accordance with this invention, the total glucosinolate levels in the seed of a hybrid Brassica plant, obtained in accordance with the teachings of the present invention, both parents of which contain less than 30  $\mu$ moles of alkenyl glucosinolates per gram dry defatted seed matter, is reduced. Preferably, hybrid plants are provided, wherein at least one of the parents has total glucosinolate levels of more than 30  $\mu$ moles, preferably more than 50  $\mu$ moles, per gram dry defatted seed and contains less than 30  $\mu$ moles of alkenyl glucosinolates per gram of defatted seed matter. Even more preferably, hybrid plants are provided, wherein at least one of the parents has total glucosinolate levels of more than 5  $\mu$ moles, preferably more than 50  $\mu$ moles, per gram in the defatted seed meal, and contains total glucosinolate levels in the whole seed basis of less than 5  $\mu$ moles per gram. Most preferably, hybrid plants are provided, from which at least one of the parents has glucosinolate levels above 0  $\mu$ moles (is the background levels found in the controls), preferably above 50 moles, has no detectable glucosinolates in the seeds (0  $\mu$ moles).

In a further embodiment of this invention, a process is provided for inhibiting expression of an s-gt gene in plant cells, particularly a process for lowering glucosinolate levels in plants, particularly Brassica plants. This process comprises the steps of transforming plant cells, preferably Brassica plant cells, with any of the above chimeric inhibitory s-gt genes, and then regenerating a plant from these transformed cells.

In yet another preferred process of the invention, plant cells are transformed with a DNA sequence encoding an antisense RNA complementary to at least part of the (m)RNA transcribed from the s-gt gene, particularly a DNA sequence encoding an antisense RNA complementary to the RNA, the cDNA of which is contained in plasmid pGL9; more particularly a DNA sequence encoding an mRNA, the cDNA of which comprises the sequence of SEQ ID No. 28 or a DNA sequence having substantial sequence similarity thereto.

Several assays are available for measuring both total and individual glucosinolates in plants or parts thereof (e.g., Quinsac and Ribailier, 1991, Assoc. Off. Anal. Chem. 74, 932-939; Reed et al., 1993, supra). Depending on the desired characteristics of the obtained plants, total glucosinolates or only glucosinolates of one type, preferably alkenyl glucosinolates, can be measured so that those plants with a reduced total glucosinolate content or with a reduced content in a certain type of glucosinolates, preferably alkenyl glucosinolates, can be selected after transformation. In accordance with this invention, the concentration of all glucosinolates formed by the S-GT enzyme, particularly the concentration of all alkenyl glucosinolates, is significantly lowered in the plants of the invention, preferentially in their seeds.

Since it is expected that other Cruciferous plants, particularly other Brassicaceae, have similar S-GT enzyme forms, differing in some characteristics but sharing regions with substantial sequence similarity, the antisense, ribozyme or co-suppression approaches outlined above can similarly be applied to lower the glucosinolate content in other Brassica species, preferably oilseed Brassica species. Preferred Brassicaceae to be transformed in accordance with this invention to inhibit s-gt expression, preferably to decrease glucosinolate content in the seed, besides Brassica napus, include Brassica juncea, Brassica oleracea, Brassica carinata, Brassica nigra, Brassica campestris and the like, and any intergenic crosses or synthetic varieties thereof.

In yet another preferred embodiment of this invention, the chim ric s-gt inhibitory gene is designed such that it effectively inhibits S-GT isoforms of both B. napus and B. juncea, so that glucosinolate production by genes of both parent genomes of these amphidiploid species is inhibited.

The thus obtained plants with lowered glucosinolate content, particularly in their seeds, are then used to cross into elite allele lines, preferably hybrid plants. Therefore, the effect of glucosinolate-reduction should preferably be apparent in the hemizygous state. In certain cases, two parent plants, transformed in accordance with this invention to have a lower glucosinolate content, particularly in their seeds, can be crossed to form a hybrid having further reduced glucosinolate content. The following Examples are offered by way of illustration and not by way of limitation. The sequence listing referred to in the Examples and the description is as follows:

- SEQ ID No. 1: B. oleraceae S-GT peptide fragment 1
- SEQ ID No. 2: B. oleraceae S-GT peptide fragment 2
- SEQ ID No. 3: B. oleraceae S-GT peptide fragment 3
- SEQ ID No. 4: B. oleraceae S-GT peptide fragment 4
- SEQ ID No. 5: B. oleraceae S-GT peptide fragment 5
- SEQ ID No. 6: B. oleraceae S-GT peptide fragment 6
- SEQ ID No. 7: B. oleraceae S-GT peptide fragment 7
- SEQ ID No. 8: primer gl3
- SEQ ID No. 9: primer gl7
- SEQ ID No. 10: 3' RACE oligo(dT) CDS primer sequence
- SEQ ID No. 11: 3' RACE Anchor primer sequence
- SEQ ID No. 12: primer gl5
- SEQ ID No. 13: primer gl9
- SEQ ID No. 14: sequence of cDNA clone pGL2-7 (incorporated primer regions have not been added, part of polyA tail is shown)
- SEQ ID No. 15: sequence of cDNA clone pGL2-25 (incorporated primer regions have not been added, part of polyA tail is shown)

- SEQ ID No. 16: primer gl1
- SEQ ID No. 17: primer glP1
- SEQ ID No. 18: primer glP2
- SEQ ID No. 19: sequence of cDNA clone pGL3-22 (incorporated primer regions have not been added)
- SEQ ID No. 20: sequence of cDNA clone pGL4-2 (incorporated primer regions have not been added)
- SEQ ID No. 21: primer glP4
- SEQ ID No. 22: primer glP5
- SEQ ID No. 23: sequence of 5' RACE Anchor (Clontech)
- SEQ ID No. 24: sequence of 5' RACE Anchor primer (Clontech)
- SEQ ID No. 25: sequence of cDNA clone pGL6-14 (incorporated primer regions have not been added)
- SEQ ID No. 26: primer gl31
- SEQ ID No. 27: primer gl32
- SEQ ID No. 28: cDNA sequence of the pGL9 clone 22 (primer regions have been included), and corresponding translated protein
- SEQ ID No. 29: amino acid sequence of the protein produced by pGL9 clone 22
- SEQ ID No. 30: cDNA sequence GT125
- SEQ ID No. 31: cDNA sequence GT135
- SEQ ID No. 32: primer 4L, inosines are represented by the code "N" at positions 3 and 18 in the sequence
- SEQ ID No. 33: primer 5R, inosines are represented by the code "N" at positions 4, 7, 13 and 19 in the sequence
- SEQ ID No. 34: sequence of s-gt genomic clone, including the coding region with intron, and leader and trailer sequences
- SEQ ID No. 35: full amino acid sequence of the S-GT protein, derived from the genomic s-gt clone

Unless otherwise stated in the Examples, all procedures for making and manipulating recombinant DNA are carried out by the standardized procedures

as described in volumes 1 and 2 of Ausubel et al., Current Protocols in Molecular Biology, Current Protocols, USA (1994) and Sambrook et al., Molecular Cloning - A Laboratory Manual, Second Ed., Cold Spring Harbor Laboratory Press, NY (1989). Standard methods and materials for plant molecular biology work are described in Plant Molecular Biology LABFAX, edited by R.R.D. Croy (1993, Bios Scientific Publishers and Blackwell Scientific publications, series editors B.D. Hames and D. Rickwood).

## **EXAMPLES**

### **Example 1**

#### **B. oleraceae S-GT protein identification**

The S-GT protein of B. oleraceae spp. botrytis was isolated as described by GrootWassink et al. (1994b, supra) which is incorporated herein by reference. The partial amino acid sequence of the isolated thiohydroximate S-glucosyltransferase was determined. Since the N-terminus was found to be blocked, the purified protein was first subjected to partial tryptic digestion. The following internal peptide fragments were obtained:

- 1) Val-Thr-Ile-Ala-Thr-Thr-Thr-Tyr-Thr-Ala-Ser-Ser-Ile-Ser-Thr-Pro-Ser-Val-Ser-Val-Glu-Pro-Ile-Ser-Asp-Gly-His-Asp-Phe-Ile-Pro (SEQ ID No. 1)
- 2) Ala-Leu-Gln-Gln-Ser-Asn-Phe-Asn-Phe-Leu-Trp-Val-Ile-Lys (SEQ ID No. 2)
- 3) Gly-His-Val-Val-Val-Leu-Pro-Tyr-Pro-Val-Gln-Gly-His-Leu-Asn-Pro-Met-Val-Gln-Phe-Ala-Lys (SEQ ID No. 3)
- 4) Ala-Thr-Leu-Ile-Gly-Pro-Met-Ile-Asp-Ser-Ala-Tyr-Leu-Asp-Lys (SEQ ID No. 4)
- 5) Leu-Pro-Glu-Gly-Phe-Val-Glu-Ala-Thr-Lys (SEQ ID No. 5)
- 6) Phe-Val-Glu-Glu-Val-Trp-Lys (SEQ ID No. 6)



- 7) Ala-Met-Ser-Glu-Gly-Gly-Ser-Ser-Asp-Arg-Ser-Ile-Asn-Glu-Phe-Val-Glu-Ser-Leu-Gly-Lys (SEQ ID No. 7)

These fragments represent about 1/4 of the B. oleraceae S-GT protein sequence and unambiguously characterize the S-GT enzyme.

Searches in protein sequence databases showed that the protein characterized by the above peptide fragments 1 to 7 is unique and has no significant sequence similarity with any previously known protein. Also, each of the peptide fragments 1 to 7 were not found to be included in any known protein. A peptide fragment of a protein showing 85 % sequence similarity with the short fragment 6 was found to be the closest match. Thus, each of the fragments identified above specifically identifies the isolated S-GT protein of B. oleraceae.

### **Example 2**

#### **Isolation of a B. napus s-gt gene**

Total RNA was prepared from B. napus cv. Westar seedlings that germinated for 5 days in the dark following routine extraction techniques. About 1 mg RNA per gram tissue was obtained as was established by OD<sub>260</sub> measurements. RNA quality was checked on a 2% TBE agarose gel (under RNase free conditions) for absence of high molecular weight (MW) DNA and integrity. mRNA was prepared from 1 mg of total RNA using the Promega kit "PolyAttract mRNA Isolation System IV" as described by the suppliers. Approximately 2-4 µg of mRNA was recovered.

From the 7 peptide sequences derived from the purified B. oleraceae S-GT protein (SEQ ID Nos 1-7), 7 sets of complementary degenerated primers (of which 2 were nested sets) were selected and synthesized by conventional methods. PCR-RACE (rapid amplifying of cDNA ends) with combining degenerate primers (based on the B. oleraceae peptide fragments) and the 3'

Anchor primer of the 3' RACE kit (Clontech) was used directly. Combinations of these primer sets were tried in PCR reactions on first strand cDNA made with the Clontech kit "3' AmpliFINDER™ RACE Kit" following the instructions of the supplier. This 3' RACE cDNA was used in PCR reactions where primers gl3 (SEQ ID No. 8) or gl7 (SEQ ID No. 9) were combined with the Anchor Primer of the Clontech 3' RACE kit (SEQ ID No. 11; complementary primer to the NN<sub>1</sub> oligo (dT)<sub>18</sub> CDS Primer (shown in SEQ ID No. 10)). Both PCR reactions were used as a template for a second semi-nested PCR with anchor primer and respectively primer ql5 (SEQ ID No. 12) or ql9 (SEQ ID No. 13). In the first case (for the gl3-gl5 combination) a PCR fragment of approximately 650bp was amplified. Upon A/T cloning of this fragment in the vector pGEM-T (Promega) following the instructions of the suppliers and electrotransformation in *E. coli* SURE (Stratagene) cells, two of the resulting transformants (pGL2 clones 7 and 25) having the appropriate insert size were sequenced. Analysis of the sequences showed an open reading frame of about 470bp, a 104bp 3'UTR and the polyA<sup>+</sup> tail for both clones. The sequence determined for pGL2-7 is shown in SEQ ID No. 14, that for pGL2-25 in SEQ ID No. 15. The amino acid sequence of the protein fragment encoded by the open reading frame contained in both these clones revealed part of the S-GT peptide 2 (as expected because this sequence was used for PCR-cloning), and the complete S-GT peptides 5, 6 (with one amino acid difference) and 7.

Using the cDNA obtained from the mRNA pool with the 5' RACE kit (Clontech) using the glP1 primer (SEQ ID No. 17) in a PCR amplification in which the degenerated primers gl1 and gl7 (SEQ ID Nos 16 and 9, respectively) were used in combination with glP2 (SEQ ID No. 18) resulted in the generation of specific PCR products of respectively 850bp or 1000bp in length. These two PCR fragments were A/T cloned in the Promega pGEM-T vector as described above and named respectively pGL4-2 and pGL3-22. The PCR product in pGL4 was completely contained in that of pGL3. pGL3 includes 982bp of s-gt ORF, of which the last 116bp overlap the s-gt fragment cloned in pGL2. The sequence of pGL3-22 is shown in SEQ ID No. 19, that of pGL4-2 in SEQ ID No. 20.

In addition to the S-GT peptide sequences 2, 5, 6 and 7 previously found in pGL2 the remaining peptide sequences 1, 3 (partially) and 4 (with one amino acid difference) were now located.

In order to clone the remaining 5' end of the B. napus s-gt gene 5' RACE was used, with a new set of nested gene specific primers located at the 5' end of the pGL3 PCR insert, namely glP4 and glP5 (SEQ ID Nos 21 and 22, respectively). The Clontech 5' AmpliFINDER™ RACE Kit was used according to the supplier's recommendations. The first strand cDNA was generated by reverse transcriptase using primer glP4, and ligated to the 5' RACE AmpliFINDER™ Anchor (SEQ ID No. 23). Upon amplification of this first strand with the kit's AmpliFINDER™ anchor primer (SEQ ID No. 24) and the gene specific primer glP5 a 300bp PCR fragment was generated. After A/T cloning (see above) and transformation in E. coli WK6 with blue-white screening, several recombinant clones were sequenced. One of them, pGL6 clone14 (pGL6-14), had the desired 5'end of the s-gt gene. The sequence showed an 88bp 5'UTR and a partial ORF of 163bp, including the full DNA sequence of S-GT peptide 3 and the start of S-GT peptide 1 (SEQ ID No. 25).

Based on the sequences of the s-gt cDNA fragments retrieved in the previous steps the forward primer gl31 (5'-TTA TTT TTC TTC JTC CTC CTC CTC T-3', SEQ ID No. 26) and reverse primer gl32 (5'-AGC AAC AAC AAC AAA CAC ACA AGA T-3', SEQ ID No. 27) were designed, located respectively in the 5'UTR and the 3'UTR. These two primers were used in PCR reactions on double-stranded (ds) cDNA of B. napus cv. Westar (Superscript Lambda System for cDNA Synthesis and -cloning, BRL, cat.no 8256RT). cDNA was amplified using 0.2 µM of these two primers in a reaction buffer consisting of 20 mM Tris-HCl (pH8.4), 50mM KCl, 2mM NaCl, 0.2 µM (each) dNTP-mixture and 2.5 Units of Taq-DNA polymerase (purchased at Gibco-BRL). Amplification conditions consisted of a initial 5 minutes denaturing at 94°C, 30 cycles of 45 sec 94°C denaturing, 60°C annealing for 45 sec, and 3 min of elongation at 72°C. As

expected, an amplification product of 1.5 kb was obtained. This DNA product was isolated and cloned in the pGEM-Z vector of Promega by A/T-cloning. One clone was retrieved, namely pGL9 clone 22 (further referred to as "the pGL9 clone"). The sequence of this clone revealed an s-gt cDNA comprising the complete open reading frame (SEQ ID No. 28 shows the full sequence obtained; the regions corresponding to the primers gl31 and gl32 are included at the 5' and 3' end, respectively (note that the region corresponding to primer gl31 is partly deleted during cloning)), with a sequence similarity highest to pGL2-25 and pGL 4-2. This cDNA comprises an open reading frame encoding a protein with a calculated m.w. of about 51 kD (50,901). The differences between the coding regions of pGL9-22 and the coding regions comprised in the partial cDNA clones obtained above (pGL2, pGL3, pGL4, and pGL6) are illustrated in Figure 2. This Figure confirms the high degree of sequence similarity between the isolated cDNA fragments. Figure 1 shows the approximate locations of the different cDNA fragments in comparison to the pGL9 cDNA clone, and the primers used in their isolation.

The cDNA clone pGL9-22, contained in the pGEM5Zf(+) vector, has been deposited in E. coli strain WK6 at the Belgian Coordinated Collections of Microorganisms (BCCM) - Laboratory for Molecular Biology - Plasmid collection (LMBP) on September 7, 1995 under accession number BCCM-LMBP 3344.

Another full length s-gt cDNA clone was obtained using the Clontech Marathon cDNA Amplification Kit. Here mRNA, isolated using the methods and kits mentioned before, was converted to double-stranded cDNA following the provided protocol. In the following amplification reaction on this material, primers gl31 and gl32 were used in combination with several heat stable DNA polymerases, such as Taq DNA polymerase (Gibco,BRL), PwoI DNA polymerase (Boehringer), Vent DNA polymerase (N w England Biolabs, Tth,XL DNA polymerase (Perkin Elmer Cetus, a commercial mixture of Tth DNA polymerase with a small amount of Vent DNA polymerase). Reaction mixtures were all made following the purchasers recommendations for optimizing conditions. Cycling

parameters after the initial 5' 94°C denaturation step were 30 cycles either of 45" 94°C, 3' 68°C or 45" 94°C, 45" 60°C, 2' 72°C. After a second round of PCR using 1 µl of a tenfold dilution of the previous PCR, both Taq DNA polymerase (conditions as mentioned above) and Tth,XL DNA polymerase (1XTth,XL DNA polymerase buffer supplemented with 1.0 mM Mg(OAc)<sub>2</sub>) showed the expected 1.5 kb amplification product. The Tth,XL DNA polymerase reaction was done using as primers gl31 and gl32 where respectively PstI and XbaI restriction sites were attached at the 5' end and 3' end. Following XbaI-PstI digestion, the PCR amplified DNA fragment was cloned in pUC19. *E. coli* MC1060 transformation gave approximately 10<sup>8</sup> transformants/µg transforming vector all having the expected insert, which was named pGL18. Partial DNA sequencing of fragments (1 fragment of 200 to 300 basepairs) of the pGL18 clone 1 shows that the pGL18-1 clone is identical to the corresponding parts in the pGL9 clone (see, e.g., Fig. 2).

The coding region of the pGL9 cDNA clone was cloned in the PstI-XbaI sites of the polylinker of the pUC19 expression vector using conventional procedures, so that it is under the control of the lac promoter. Expression of this chimeric gene in *E. coli* by IPTG induction showed a band recognized by rabbit polyclonal anti-*B. oleraceae* S-GT antibody, while this band was not present in the untransformed induced strain, upon Western blotting. The recombinantly produced S-GT protein was confirmed to have a glucosyl-S-transferase activity in the radioassay of GrootWassink et al. (1994b, supra). This assay was based on the incorporation of [<sup>14</sup>C]Glc from UDP-Glc into phenylacetothiohydroximate. As starting material, the above *E. coli* strains containing the pGL14 clone were used. The wild type WK6 *E. coli* strain was used as a control, as well as water. Strains were grown and induced (with IPTG) overnight. A Western blot and a Coomassie staining were first performed to check for expression of the formed fusion protein. The protein concentration was assessed by a Bradford assay, using BSA as a standard. For the glucosinolate assay, 80 µl of reaction mixture (50mM Mes buffer with UDP-[U<sup>14</sup>C]Glc (0.05 µCi) and 1.0 mM phenyl acetothiohydroximate, 5 mM MgSO<sub>4</sub> and 0.1% ME) was mixed with 20 µl of

extract (from 1 ml of overnight culture that went 3 times through the French press, 800 Pa/inch<sup>2</sup>; in a 1/100 or 1/500 dilution, depending on the protein concentration of each of the two samples). The samples were incubated at 30°C in a water bath and the reaction was stopped by boiling for 10 min. After cooling on ice, 500 µl ethylacetate was added and the tube containing the samples were vigorously shaken. Samples were spun down and the top phase (containing the glucosinolate in ethylacetate) was recovered for radioactivity measurement.

The results showed that, at the proper dilution of the cell extract (between 1/100 and 1/500 depending on the sample), the enzyme activity increased with time and was significantly higher than the controls. Using the 0 min reaction as a control, the following relative enzyme activities were obtained after different time periods: 22 after 10 minutes (i.e., the glucosyltransferase activity after 10 minutes of incubation with the reaction mixture was 22 times that of the activity of the recombinant s-gt strain at time 0), 30 after 20 min, 40 after 30 min and 80 after 60 min.

Using the wild type E. coli strain as a control, the following relative enzyme activities were observed for the S-GT-producing clone: 1 after 0 min reaction, 23 after 10 min (i.e., after 10 minutes of reaction, the S-GT producing strain had 23 times the glucosyl-transferase activity of the wild type strain), 16 after 20 min, and 65 after 30 min. Thus, these radioassays show that the s-gt gene cloned in the transformed E. coli WK6 produced a protein with a significant S-GT activity, thus confirming that the correct DNA sequence had been isolated. The S-GT proteins encoded by the full open reading frames of the other isoforms corresponding to the cDNA fragments isolated above show the same significant S-GT activity under similar assay conditions.

The same S-GT assay was done with leaf material obtained from B. napus varieties Jet Neuf, Express and Vivol. Of these varieties, Jet Neuf is known to be a high glucosinolat variety and Express and Vivol are known to be low glucosinolate varieties.

For this assay, leaf material was randomly taken from several plants of the B. napus lines Vivol, Express and JetN uf. The sampling took place 110 days after sowing and after a normal vernalisation period, shortly before flowering. Leaf samples were quickfrozen in liquid N<sub>2</sub> immediately after picking and stored at -70°C until further processing. Leaf material was grinded in liquid N<sub>2</sub> using a mortar and pestle. An equal volume/weight of extraction buffer (0.2 M Hepes pH7.5; 5mM EDTA; 0.1%  $\beta$ -MeEtOH) was added to the powder and shaken vigorously for 30 minutes at 4°C. The supernatant of these leaf extracts was concentrated approximately 10-fold using centrprep-30 concentrators (Amicon). Protein concentrations were measured with the Bradford method:

B.napus      Vivol: 52.8 mg/ml  
                  Express: 66.0 mg/ml  
                  JetNeuf: 37.4 mg/ml

The S-GT radio assay was performed as described in GrootWassink et al. (1994b, supra). As controls were used: TE-buffer (negative control) and an E.coli extract from the WK6(pGL14) strain (s-gt cDNA ORF in frame with LacZ and under control of the Lac-promoter, in plasmid pUC19). Averages of repeated scintillation countings were used in the calculations of the specific activities ( $\Delta$ cpm/sec/mg protein,  $\Delta$ cpm being the difference in amount of cpm at a certain time point and at time 0'). This assay was performed on the extracts and dilutions thereof at different time points (0', 5', 10', 20', 30', 1h, 2h, 4h), to determine optimal conditions to perform the assay on these leaf extracts.

The table below shows the specific activities ( $\Delta$ cpm/sec/mg protein) of the leaf extracts of B. napus Vivol (10x dilution), Express (10x dilution), JetNeuf (20x dilution) on 0 minutes and 20 minutes.

time	Vivol	Express	JetNeuf
0'	0	0	0
20'	2.62	1.65	19.04

The obtained data clearly show an approximate 10-fold higher specific activity of the high glucosinolate containing B. napus JetNeuf line compared with the two other low glucosinolate lines. Other batches of the above plant material gave the same results, showing that indeed the high glucosinolate varieties have a higher specific activity of S-GT in their tissues. This strongly indicates that decreasing the S-GT activity in the plants by expression of an s-gt inhibitory gene will produce low glucosinolate plants.

Furthermore, several other cDNA fragments encoding S-GT or parts thereof were isolated from a cDNA library (ZipLox, GibcoBRL, Life Technologies, INC) of B. napus cv. Westar. Lambda ZipLox (GibcoBRL Life Technologies Inc.) is a Lambda expression vector that combines cDNA cloning and screening of Lambda libraries. The cDNA can be recovered in an autonomously-replicating plasmid using an in vivo excision protocol.

About five million clones were screened in total. Nylon filters (Hybond<sup>TM</sup>-N, Amersham Life Science) were left in contact with the agarose plate for 1 min, then denatured (0.5 M NaOH/1.5 M NaCl) for 5 min, neutralized (1.5 M NaCl/0.5 M TrisHCl, pH 7.5) for 5 min and rinsed with 2X SSC for 10 min with gentle shaking. The filters were air dried for 10 min and baked for 2 hrs at 65°C. Two amplified fragments obtained by the RACE procedure (pGL2-7 and pGL3-22, SEQ ID No. 14 and 19, respectively) were used as probes. Inserts were recovered from the recombinant plasmids by double enzymatic digests (SacII and PstI) and were about 500 bp (clone pGL2-7) and 1 kb (clone pGL3-22) long. Probes were radio-labeled with [<sup>32</sup>P] dCTP using a random priming kit (GibcoBRL, Life Technologies, Inc.), purified with a Nick<sup>TM</sup> column (Sephadex<sup>R</sup> G-50 DNA grade, Pharmacia Biotech), boiled for 10 min and ice-chilled before use. The specific activity of the probe was checked by scintillation counting (liquid scintillation counter 1219 Rackbeta, LKB Wallac).



Filters were pre-hybridized overnight with 15ml of "SureHyb" (5X Denhardt's solution, 5X SSC, 2% Lauryl Sarcosine and 10% Dextran sulphate) solution at 65°C. Hybridizations were performed overnight at 65°C with the same solution containing the labeled probe (approximately 1-20 ng). Filters were washed with 2X SSC/0.1% SDS for 20 min at 65°C, and with 1X SSC/0.1% SDS for 20 min at 65°C, and then exposed overnight to X-ray film (Kodak XAR 5) with a "hi-speed" intensifying screen at -80°C.

Three rounds of purification were performed on each positive clone. When single positive colonies could be singled out they were picked and incubated in 500 µl SM buffer (Sambrook et al., 1989, Molecular Cloning. A laboratory manual. 2nd edition. Cold Spring Harbor Laboratory press. Cold Spring Harbor, New York) at 4°C before plasmid DNA extraction.

Recombinant DNA was isolated using either the plasmid excision approach provided by the ZipLox system or the Lambda DNA extraction using the phage burst method as described by Sambrook et al. (1989). For the direct excision of the plasmid, 0.5 µl of the Lambda ZipLox phage in 1 ml of SM buffer were incubated with 200 µl of *E. coli* DH10B (overnight culture in Luria-Bertani (LB) medium supplemented with 0.2% w/v maltose and 10 mM magnesium) for 60 min at 37°C. Aliquots of 50 µl of this mixture were plated on LB solid medium, supplemented with 0.2% w/v maltose, 10 mM magnesium and 100 µg/µl ampicillin. Single large colonies were grown overnight in liquid LB medium (with ampicillin) prior to DNA extraction with the Plasmid Maxi Prep (QiaGen). Both strands of the plasmid DNA from each positive clone were sequenced using the Taq Dideoxy<sup>TM</sup> Terminator cycle sequencing kit (Applied Biosystems) on a 370A sequencer (Applied Biosystem).

Four positive clones have been identified and sequenced. The first clone (GT 177) covers the last third of the gene at the 3'-end, and has 94 % DNA

sequence similarity with the pGL9 sequence. Three other clones (GT 135, GT 124 and GT 125) have both strands fully sequenced (see SEQ ID No. 30 and 31 for DNA sequence of GT 125 and 135 respectively). Their sequences also had a high degree of sequence similarity with pGL9, as well as with the genomic clone (see further). The inferred amino acid sequences contained those of the seven peptides (SEQ ID Nos 1-7), except for GT 125, which was found to have differences at the 5' side. These differences in GT 125 are probably an artifact caused by a recombination event. In GT 135, a more upstream stop codon was present so that a shorter protein is encoded. Clones GT 124 and GT 135 are very similar in length (1582 and 1585 nucleotides respectively) and in sequence. These additional cDNA sequences, although not representing full coding sequences, confirm the existence of isoforms of the s-gt gene observed in the previous cDNA isolations and show that these isoforms have significant sequence similarity. The GT 135 cDNA sequence was most similar to the pGL6 cDNA sequence, while the GT125 cDNA sequence was most similar to the pGL9 cDNA sequence.

Genomic DNA analysis of several Brassica napus cultivars indicated that at least two s-gt genes are present in B. napus. Genomic DNA was extracted from B. napus cultivars Cresor, Cyclone and line 94COO3008. Cresor has a high level of glucosinolates, Cyclone is a low-glucosinolate variety (6  $\mu$ moles per gram dry defatted seed meal) and line 94COO3008 is an intermediate type (48  $\mu$ moles per gram dry defatted seed meal).

Four restriction enzymes (EcoRI, BamHI, XbaI and HindIII, Pharmacia Biotech) were used to digest 10  $\mu$ g of Brassica genomic DNA. DNA fragments were resolved by electrophoresis including a DNA molecular size marker (BRL). After transfer onto nylon membrane (Hybond N+, Amersham Life Science) by vacuum transfer (Tyler), hybridizations were performed with labeled probes (fragments PGT2#7 or PGT3#22, respectively SEQ ID Nos 14 and 19). The obtained banding patterns revealed that there are at least two S-GT genes

present in B. napus but probably no more. The cDNA sequence analysis already suggested the presence of two slightly different s-gt DNA sequences in B. napus.

These two different genes are thought to come from one of the progenitors of B. napus, i.e., B. rapa and B. oleraceae. No polymorphism could be detected within the three B. napus cultivars (high, low and intermediate levels of glucosinolates), suggesting that the cause of the difference in glucosinolate content among B. napus species does not reside within the structure of the S-GT gene.

Finally, the complete Brassica napus s-gt gene was cloned from a genomic DNA library using the above obtained sequence information. A B. napus cv. Bridger genomic EMBL3 Bacteriophage DNA library (Karn et al., 1980, Proc. Natl. Acad. Sci. USA 77:5172-5176) was obtained from Clontech and was used to isolate a recombinant DNA phage which contained a nucleotide sequence encoding the S-GT protein. Degenerate oligonucleotides 4L (SEQ ID No. 32, CCIATGATHGAYAGYGCITAY (Y, H are IUPAC codes for degenerate positions (Y: C or T, H: A or C or T)) and 5R (SEQ ID No. 33, MAAIGTIGCYTCIACRAAICCYTC, M is IUPAC code for A or C, R is IUPAC code for A or G at this position)) were synthesized using a Beckman Oligo 1000M DNA Synthesizer and by following the manufacturers protocols. Using these degenerate primers, a short DNA fragment was amplified and used as a probe to screen the genomic library. The putative  $\lambda$ clone BnGT was isolated. From this clone a short 6 kb Sall DNA fragment was subcloned into the plasmid vector pUC19 (Messing, 1983, Methods Enzymol. 101:20-78) using standard techniques (1989, Molecular Cloning: A laboratory Manual. Cold Spring Harbor Laboratory Press), thus creating pGT6Sal.

Nucleotide sequence analysis of this 6 Kb Sall DNA fragment was performed using a Perkin Elmer ABI 373 DNA Stretch Sequencer. Double stranded plasmid DNA was used as the template using the ABI PRISM Dye Terminator Cycle Sequencing Kit and by following the supplied protocols.

Detailed DNA sequence information of the protein-encoding part and some leader and trailer-sequences is shown in SEQ ID No. 34. This sequence is 1969 bp in length and contains the DNA sequence which directs the synthesis of the protein thiohydroximate s-glucosyltransferase. SEQ ID No. 35 shows the predicted amino acid sequence of the complete enzyme. The expected start site of this coding region is at position 213 of the listed DNA sequence. The coding sequence is interrupted once by an intron which is 173 bp long and starts at position 882 of the listed sequence and ends at position 1054. The coding sequence is estimated to be 1570 bp long and terminates at position 1783 as listed in SEQ ID No. 34. The complete encoded protein is 466 amino acids in length and has an estimated nonglycosylated molecular weight of 51.3 kD. The genomic clone has been deposited at the ATTC on October 31, 1996. Sequence comparison of this genomic sequence with the pGL9 cDNA sequence obtained above shows that there is very high homology between the pGL9 cDNA and this genomic sequence, so that the antisense strategy will undoubtedly lead to inactivation of all the endogenous s-gt genes. Also, comparison of the other cDNAs obtained above show that there were essentially two groups of sequences: one group formed by the s-gt genomic clone, pGL9, GT125 and another group formed by GT135 and pGL6. In these groups the sequence similarity is considerably higher than between the groups, again confirming the presence of two isoform genes in B. napus.

The promoter region of the s-gt gene is identified in the genomic clone corresponding to the cDNA contained in plasmid pGL9 (as deposited under deposit number BCCM-LMBP 3344). The promoter region, consisting of a DNA sequence comprising 1000 bp upstream of the open reading frame, is isolated for further construction work. In the context of the present invention, this region is particularly isolated from the genomic clone designated pGT6Sal deposited at the ATCC on October 31, 1996. The promoter region and the leader sequence are characterized by the partial nucleotide sequence upstream of the coding region shown in SEQ ID No. 35 from position 1 to position 212.

**Example 3****Antisense inhibition of s-gt gene expression in oilseed rape plants**

Using standard plant molecular biology techniques, several constructs based on the above isolated full cDNA (pGL9, SEQ ID No. 28) are inserted into oilseed rape plant cells which are then regenerated into plants. A B. napus variety having more than 30  $\mu$ moles of alkenyl glucosinolate per gram oil free seed matter is transformed with the s-gt antisense constructs of Example 3. Hypocotyl explants of Brassica napus are obtained, cultured and transformed essentially as described by De Block et al. (1989, Plant Physiol. 91, 694), except for the following modifications:

- hypocotyl explants are precultured for 3 days in A2 medium [MS, 0.5 g/l Mes (pH 5.7), 1.2% glucose, 0.5% agarose, 1 mg/l 2,4-D, 0.25 mg/l naphthalene acetic acid (NAA) and 1 mg/l 6-benzylaminopurine (BAP)].
- infection medium A3 is MS, 0.5 g/l Mes (pH 5.7), 1.2% glucose, 0.1 mg/l NAA, 0.75 mg/l BAP and 0.01 mg/l giberellinic acid (GA3).
- selection medium A5 is MS, 0.5 g/l Mes (pH 5.7), 1.2% glucose, 40 mg/l adenine.SO<sub>4</sub>, 0.5 g/l polyvinylpolypyrrolidone (PVP), 0.5% agarose, 0.1 mg/l NAA, 0.75 mg/l BAP, 0.01 mg/l GA3, 250 mg/l carbenicillin, 250 mg/l triacillin, 0.5 mg/l AgNO<sub>3</sub>.
- regeneration medium A6 is MS, 0.5 g/l Mes (pH 5.7), 2% sucrose, 40 mg/l adenine.SO<sub>4</sub>, 0.5 g/l PVP, 0.5% agarose, 0.0025 mg/l BAP and 250 mg/l triacillin.
- healthy shoots are transferred to rooting medium which was A8: 100-130 ml half concentrated MS, 1% sucrose (pH 5.0), 1 mg/l isobutyric acid (IBA), 100 mg/l triacillin added to 300 ml perlite (final pH6.2) in 1 liter vessels (MS stands for Murashige and Skoog medium (Murashige and Skoog, 1962, Physiol. Plant. 15, 473)).

Hypocotyl explants are infected with Agrobacterium tumefaciens strain C58C1Rif<sup>R</sup> carrying:

- a helper Ti-plasmid pMP90 (Koncz and Schell (1986), Mol. Gen. Genet. 204, 383) or a derivative thereof (such as pGV4000), which is obtained by insertion of a bacterial chloramphenicol resistance gene linked to a 2.5 kb fragment having similarity with the T-DNA vector pGSV8, into pMP90.
- T-DNA vector pGSV8 containing between the T-DNA borders the chimeric s-gt inhibitory gene.

B. napus plants are transformed to contain the following chimeric s-gt inhibitory gene in plant transformation vector pGSV8: a chimeric gene comprising:

- the 35S3 promoter (Hull and Howell, 1987, Virology 86, 482-493) operably linked to one of the following different DNA sequences:

- 1) a DNA encoding an antisense RNA complementary to the about 0.7 Kb Styl-XhoI fragment of the s-gt DNA sequence contained in pGL9 (the 5' half of the s-gt coding region, in vector pTKV9) or
- 2) a DNA encoding an antisense RNA complementary to the about 0.65 Kb Avall-AsnI fragment of the s-gt DNA sequence contained in pGL9 (the 3' half of the s-gt coding region) (in vector pTKV10), or
- 3) a DNA encoding an antisense RNA complementary to the about 1.3 kb Styl-AsnI fragment of the s-gt DNA sequence contained in pGL9 (the almost complete s-gt coding region in vector pTKV8).

These 3 sequences have each been operably linked to the 3' transcript termination and polyadenylation region of the nopaline synthase gene (DePicker et al., 1982, J. Mol. Appl. Gen. t. 1, 561), thus giving 3 chimeric genes, each comprising a DNA encoding a different antisense RNA.

To select the transformed cells, a chimeric bar gene (De Almeida et al., 1989, Mol. Gen. Genet. 218, 78) conveying resistance to phosphinotricin was included in the transforming DNA. The bar coding region (Thompson et al., 1987, The EMBO J. 6, 2519) is under the control of the promoter of the Arabidopsis thaliana ribulose-1,5-biphosphate carboxylase small subunit 1A gene (Krebbers et al., 1988, Plant Mol. Biol. 11, 745) and is flanked by the 3' transcript termination and polyadenylation region of gene 7 (Velten and Schell, 1985, Nucleic Acids Res. 13, 6981-6999).

The thus obtained vectors pTKV8, 9 and 10, carry the selectable marker gene in two orientations compared to the s-gt antisense DNA chimeric gene, yielding pTKV8a, pTKV8b, pTKV9a, pTKV9b, pTKV10a, and pTKV10b (a in same orientation, b in opposite orientation of transcription). Plasmid pTKV8a, producing an antisense RNA in plant cells complementary to most of the pGL9 s-gt coding region, was deposited in host cell E. coli MC1061 at the BCCM-LMBP on September 7, 1995 under accession number LMBP 3343.

These vectors pTKV8a, pTKV9a, and pTKV10a have also been used to transform several B. napus and B. rapa varieties by the above-mentioned transformation protocol using Agrobacterium tumefaciens. For the group of low glucosinolate B. napus cultivars, cv. Vivol was transformed with these three vectors. From the group of high glucosinolate B. napus cultivars, cv. JetNeuf was transformed with these three vectors. Furthermore, also the low-glucosinolate B. napus cultivar Express is transformed with these three vectors. Of B. rapa, plant cells of the high-glucosinolate cultivars Bele and Tyko have been transformed with the pTKV8a vector and they are regenerated into plants. Of these transformations, Southern analysis of rooted plantlets confirm the presence of the constructs in the plant's genome. After being selected and sent to the greenhouse the winter oilseed rape lines enter a vernalization of two months in order to flower afterwards. At the growth-stage 2.8 (maning 8 expanded true leaves), 3 leaves are picked and quickfrozen in liquid nitrogen. These samples are analyzed for the s-gt antisense RNA by Northern blotting, for the S-GT

enzyme activity by the above-mentioned enzymatic assay, and their glucosinolate patterns are analyzed by means of HPLC. The lines giving best reduction in glucosinolate levels in the seeds, correlated with the expression of the antisense construct in the plant are selected for further breeding.

In addition to the s-gt antisense constructs described above, also vectors based on the full genomic sequence, including the intron sequence, are made to complete the antisense strategy.

Following similar procedures, the above (antisense) plant transformation constructs are also made with DNA encoding antisense RNA complementary to the RNA corresponding to all or part of the open reading frame corresponding to the pGL6 clone described above. In other transformation steps, Brassica juncea varieties are also transformed with the pGL9-derived antisense constructs pTKV8, pTKV9 and pTKV10, so as to obtain plants with markedly decreased glucosinolate content in their seeds.

Further, some constructs comprising the S-GT 1000 bp upstream promoter sequence instead of the CaMV 35S promoter are made. Because of the expression pattern of the endogenous promoter, the endogenous s-gt promoter is a particularly preferred embodiment of this invention. Indeed, it is thought that the s-gt promoter will express the antisense construct in those cells where transcription of the endogenous s-gt gene is most active, thus most effectively targetting the desired sites.

#### **Example 4**

##### **Plant selection and agronomic evaluation**

The transformed oilseed lines of Example 3, having a single copy antisense s-gt chimeric gene and having reduced glucosinolate content in the seeds, are selected upon transformation. These selected transformed plants are rendered homozygous by doubled haploid methodology, and the obtained



homozygous lines are then crossed with a male sterile B. napus variety (obtained by following the teachings of EP 344029 and Mariani et al. (1990, Natur 347, 737-741; 1992, Nature 357, 384-387)) to obtain a hybrid having reduced glucosinolate content and being male sterile. The decrease of glucosinolate content in hybrids obtained from crosses of different B. napus varieties with known glucosinolate levels with the obtained selected male sterile hybrid plants allows the quantitative determination of the reduction in glucosinolate content in a hemizygous state in a hybrid plant. A statistically significant reduction (see, for example Statistical Methods. eds. Snedecor. G.W. and Cochran. Wig. Iowa State, University Press Iowa, USA, 1976) in alkenyl glucosinolate levels is found in the seeds of selected hybrid plants, thus illustrating that the glucosinolate content in Brassica varieties still having high glucosinolate levels in their seeds can be lowered to agronomically acceptable levels.

Agronomic evaluation of the obtained transformed plants and the resulting hybrids shows that these plants have good yields when compared to hybrids obtained from crossing plants not transformed with the s-gt antisense chimeric gene.

Additional plants wherein the invention is applicable are all other plants wherein a reduced expression of an UDP-glucose:thiohydroximate S-glucosyltransferase is desired, or wherein reduction of glucosinolate content is desired in view of the negative effects of the glucosinolates on animal or human consumption. Exemplary crops wherein the above embodiments are as well applicable are other oilseed rape species and other Crucifer crops, particularly other Brassica crops, including but not limited to: B. oleraceae, B. campestris, B. rapa, B. juncea, and B. nigra. Because of the large sequence similarity between the different S-GT isoforms and their encoding genes, it is thought that the constructs of Example 3 can equally well be used to obtain a reduced glucosinolate content in the other Brassica plants, preferably in their seeds.

The examples and embodiments of this invention described herein are only supplied for illustrative purposes. Many variations and modifications in accordance with the present invention are known to the person skilled in the art and are included in this invention and the scope of the claims. For instance, it is possible to alter, delete or add some nucleotides or amino acids to the DNA and protein sequences of the invention without departing from the essence of the invention.

All publications (including patent publications) referred to in this application are hereby incorporated by reference.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

## (i) APPLICANT:

(A) NAME: PLANT GENETIC SYSTEMS N.V.  
 (B) STREET: Plateaustraat 22  
 (C) CITY: Gent  
 (E) COUNTRY: BELGIUM  
 (F) POSTAL CODE (ZIP): B-9000  
 (G) TELEPHONE: 32 9 2358454  
 (H) TELEFAX: 32 9 2240694

(A) NAME: NATIONAL RESEARCH COUNCIL OF CANADA  
 (B) STREET: Montreal Road  
 (C) CITY: Ottawa  
 (D) STATE: Ontario  
 (E) COUNTRY: CANADA  
 (F) POSTAL CODE (ZIP): K1A 0R6  
 (G) TELEPHONE: 1 613 9933899  
 (H) TELEFAX: 1 613 9526082

(ii) TITLE OF INVENTION: Plants with reduced glucosinolate content

(iii) NUMBER OF SEQUENCES: 35

## (iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
 (B) COMPUTER: IBM PC compatible  
 (C) OPERATING SYSTEM: PC-DOS/MS-DOS  
 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

## (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Val	Thr	Ile	Ala	Thr	Thr	Thr	Tyr	Thr	Ala	Ser	Ser	Ile	Ser	Thr	Pro
1				5				10						15	
Ser	Val	Ser	Val	Glu	Pro	Ile	Ser	Asp	Gly	His	Asp	Phe	Ile	Pro	
			20					25					30		

## (2) INFORMATION FOR SEQ ID NO: 2:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Ala	Leu	Gln	Gln	Ser	Asn	Phe	Asn	Phe	Leu	Trp	Val	Ile	Lys
1				5					10				

## (2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 22 amino acids  
    (B) TYPE: amino acid  
    (C) STRANDEDNESS:  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Gly His Val Val Val Leu Pro Tyr Pro Val Gln Gly His Leu Asn Pro  
1                    5                    10                    15  
Met Val Gln Phe Ala Lys  
                    20

## (2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 15 amino acids  
    (B) TYPE: amino acid  
    (C) STRANDEDNESS:  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Ala Thr Leu Ile Gly Pro Met Ile Asp Ser Ala Tyr Leu Asp Lys  
1                    5                    10                    15

## (2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 10 amino acids  
    (B) TYPE: amino acid  
    (C) STRANDEDNESS:  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Leu Pro Glu Gly Phe Val Glu Ala Thr Lys  
1                    5                    10

## (2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 7 amino acids  
    (B) TYPE: amino acid  
    (C) STRANDEDNESS:  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Phe Val Glu Glu Val Trp Lys  
1                    5

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Ala Met Ser Glu Gly Gly Ser Ser Asp Arg Ser Ile Asn Glu Phe Val  
1 5 10 15  
Glu Ser Leu Gly Lys  
20

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

CARGARWSNA AYT TYAAYTT

20

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CCNTAYCCNG TNCARGGNCA

20

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CCTCTGAAGG TTCCAGAATC GATAGGAATT CTTTTTTTTT TTTTTTTTTTV N

51

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- PHYSICAL CHARACTERISTICS:
- (A) LENGTH: 38 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CTGGTTCGGC CCACCTCTGA AGGTTCCAGA ATCGATAG 38

(2) INFORMATION FOR SEQ ID NO: 12:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

AAATTTYYTNT GGGTNATHAA 20

(2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

CCNATGGTNC ARTTYGCNAA 20

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 566 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

AGAAGCTCAT ATAGCGAAGT TACCAGAAGG GTTTGTGGAA GCTACCAAAG ACAGAGCGTT 60  
 GCTTGTTTCT TGGTGTAACC AGCTTGAGGT TTAGCTCAT GGATCTATAG GTTGTTTTTT 120  
 GACTCACTGC GGTGGAAC TCGACGCTGGA AGGGTTGAGT TTGGGAGTTC CGATGGTGGG 180  
 TGTGCCGCAG TGGAGTGATC AGATGAATGA TGCTAAGTTT GTGGAGGAGG TTTGGAGAGT 240  
 TGGGTATAGA GCGAAAGATG AAGCTGGGGG AGGAGTTGTG AAGAGCGATG AGGTGGTGAG 300  
 GTGTTTGAAA GGAGTGATGG AAGGAGAGAG TAGTGTGGAG ATTAGAGAAA GTTCTAAGAA 360  
 ATGGAAAGAT TTGGCTGTGA AGGCGATGAG TGAAGGAGGA AGCTCTGATC GGAGCATTAA 420  
 TGAGTTTGTT GAGAGTTTAG GGAAGAAACA TTGAGAGGTA ATGAGATTG TAAATCTTGT 480  
 GTGTTTGTTG TTGTTGCTCA AGAGCATTGT ACGGAGCGGA TTGTCATTCA GTAATATGAA 540  
 TAAACCAATT GTGATAGTAA AAAAAA 566

(2) INFORMATION FOR SEQ ID NO: 15:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 568 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

```

AGAAGCTCAT ATAGCGAAGT TACCAGAAGG GTTTGTGGAA GCTACCAAAG ACAGAGCGTT      60
CCTTCCTTTCT TCGTCTAACC ACCTTCAGGT TTTACCTCAT CAATCCATAG GTTGCTTTTT      120
GACTCACTGC GGTTGGAAGT CGACGTTGGA AGGATTGAGT TTGGGAGTTC CGATGGTTGG      180
TGTGCCTCAG TGGAGTGATC AGATGAATGA TGCTAAGTTT GTGGAGGAGG TTTGGAGAGT      240
TGGGTATAGA GCGAAGGAGG AAGCTGGGGG AGGAGTTGTG AAGAGCGATG AGGTGGTGAG      300
GTGTTTGAGA GGAGTGATGG AAGGAGAGAG TAGTGTGGAG ATTAGAGAGA GTTCTAAGAA      360
GTGGAAGAT TTGGCTGTGA AGGCGATGAG TGAAGGAGGA AGCTCTGATC GGAGCATTAA      420
TGAGTTTGTG GAGAGTCTAG GGAAGAAACA TTGAGAGGTA ATGAGATTTG TAAATCTTGT      480
GTGTTTGTTG TTGTTGCTCA AGAGCATTGT ACGGAGCGGA TTGTCATTCA GTAATATGAG      540
TAAACCAATT GTGATATTTG AAAAAAAAAA      568

```

(2) INFORMATION FOR SEQ ID NO: 16:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

```

GAYGGNCAYG AYTTYATHCC      20

```

(2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

```

CCTTCCAGCG TCGAGTTCCA ACCGC      25

```

(2) INFORMATION FOR SEQ ID NO: 18:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

CGGAATTCGC TAAACCTCA AGCTGGTTAC ACC

33

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 940 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

CACCTCAACC CAATGGTCCA GTTCGCTAAA CGCCTAGTCT CCAAAGGCCT CAAAGTCACA	60
ATCGCCACCA CCACCTACAC CGCCTCCTCC ATCTCCACCC CCTCCGTCTC CGTCGAACCA	120
ATCTCCGACG GCCACGACTT CATCCCCATA GCGTCCCCCG GCGTCAGCAT CGACGCCTAC	180
TCCGAATCCT TCAAGCTCCA CGGCTCCCAA ACCTTAACCC GCGTAATCTC CAAATTCAAA	240
TCCACAGATT CCCCCATCGA TTCTTTAGTC TACGACTCTT TCCTCCCGTG GGGACTCGAA	300
GTCGCGAGAT CCAACTCCCT CTCAGCTGCC GCTTTCTTCA CCAACAACCT CACCGTTTGC	360
TCTGTCCTTC GCAAATTCGC CTCCGGTGAG TTTCTCTCC CCGCTGATCC CGCTTCCGCG	420
CTGTATCTCG TCCGTGGCTT GCCGGCTTTG AGCTACGACG AGCTTCCTTC CTTCGTGGGC	480
CGTCACTCGT CGAGCCACGC CGAACACGGG AGAGTTCTTC TGAACCAGTT CATTAAACCAT	540
GAAGATGCTG ATTGGCTGTT CGTCAACGGC TTCGAAGGGT TAGAGACACA AGGTTGTGAA	600
GTTGGAGAAT CAGAGACTAT GAAGGCGACG TTGATCGGAC CTATGATCCC ATCTGCTTAT	660
CTTGACGCCC GAATCAAAGA CGATAAAGGC TACGGCTCGA GTCTGATGAA GCCGCTCTCG	720
GAGGAGTGTA TGGAGTGGTT AGACACTAAG CTGAGTAAGT CGGTGGTTTT TGTTCGTTT	780
GGTTCCTTTA GGATCCTCTT TGAGAAGCAA CTAGCTGAGG TAGCAACGGC GTTACAAGAA	840
TCCAACTTTA ACTTCTTG TGATGATGTA GAAGCTCGTA TAGCGAAGTT ACCAGAAGGG	900
TTTGTGGAAG CTACCAAAGA CAGAGCGTTG CTTGTTTCTT	940

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 794 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO



(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

CATAGGCGTT CCCGGCGTCA GCATCGACGC ATACTCCGGA TCCTTCAAGC TCAACGGCTC	60
CGAAACCTTA ACCCGAGTAA TCTCAAAATT CAAATCCACA GATTCACCCA TCGATTCAAT	120
AGTCTACGAC TCTTTCCTCC CGTGGGGACT CGAAGTCGCG AGATCTAACT CCATCTCAGC	180
TGCTGCTTTC TTCACCAACA ACCTCACCGT TTGCTCTGTT CTACGCAAAT TCGTCTCCGG	240
TGAGTTTCCT CTCCCCGCTG ATCCCGCTTC CGCGCCGTAT CTCGTCCGTG GCTTACCGGC	300
TTTGAGCTAC GACGAGCTTC CTTCTTCGT CGGACGTCAC TCGTCGAGCC ACGCGGAGCA	360
CGGGAGAGTT CTTCTGAACC AGTCCGTAA CCACGAAGAT GCTGATTGGC TGTTCGTCAA	420
CGGCTTCGAA GCCTTAGACA CACACCTTC TCACTTGA GAATCAGAGG CGATGAAGGC	480
GACGTTGATC GGACCTATGA TACCATCTGC TTATCTCGAC GGCCGAATCA AAGACGATAA	540
AGGCTACGGC TCGAGCCTGA TGAAGCCGCT CTCGGAGGAG TGTATGGAGT GGTTAGACAC	600
TAAGCTGAGC AAGTCGGTGG TTTTGTTC GTTTGGTTCC TTTGGGATCC TCTTTGAGAA	660
GCAACTCGCT GAGGTGGCAA AGGCGTTACA AGAATCCAAC TTAACTTCT TGTGGGTGAT	720
CAAAGAAGCT CATATAGCGA AGTTACGAGA AGGGTTTGTG GAAGCTACCA AAGACAGAGC	780
GTTGCTTGTT TCTT	794

(2) INFORMATION FOR SEQ ID NO: 21:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 26 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

AGACGGAGGG GGTGGAGATG GAGGAG	26
------------------------------	----

(2) INFORMATION FOR SEQ ID NO: 22:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 35 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

CGGAATTCGG TGTAGGTGGT GGTGGCGATT GTGAC	35
--	----

(2) INFORMATION FOR SEQ ID NO: 23:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 35 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CACGAATTCA CTATCGATTG TGGAACCTTC AGAGG

35

(2) INFORMATION FOR SEQ ID NO: 24:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 38 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

CTCTTCCCG CACCTCTCA ACCTTCGCA ATCGATAC

36

(2) INFORMATION FOR SEQ ID NO: 25:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 223 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

GATCAGACTT ATTATTTTTC TTCTTCCTCC TCCTCTTCTC AGTCTTCTTC AACTGAAAAC 60  
AAACAGAAAC TAAGGCTTCA AAGTCACAAT GGTGGAAACA ACAACAACAA CAACAGCAAA 120  
GACCAGCTCC AAAGGCCACG TCTTGGTCTT ACCTTACCCA GTCCAAGGCC ACCTCAACCC 180  
AATGGTCCAG TTCGCTAAAC GCCTAGTCTC CAAAGGCCTC AAA 223

(2) INFORMATION FOR SEQ ID NO: 26:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

TTATTTTCTTCT TCCTTCCTCT CCTCT

25

(2) INFORMATION FOR SEQ ID NO: 27:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

AGCAACAACA ACAAACACAC AAGAT

25

## (2) INFORMATION FOR SEQ ID NO: 28:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1513 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 69..1466

## (xii) SEQUENCE DESCRIPTION. SEQ ID NO. 28.

```

TTATTTTCT TCTTCCTCCT CCTCTGTGTC TTTGTCAACT GCAAACAAA CAGAAACCAA      60
GTTCTGCA ATG GCG GAA ACA ACA ACA ACA ACA GCG ACC AAC TCC AAA      110
  Met Ala Glu Thr Thr Thr Thr Thr Thr Ala Thr Asn Ser Lys
      1              5              10

GGC CAC GTC GTG GTC TTA CCT TAC CCA GTC CAA GGC CAC CTC AAC CCA      158
Gly His Val Val Val Leu Pro Tyr Pro Val Gln Gly His Leu Asn Pro
  15              20              25              30

ATG GTC CAG TTC GCT AAA CGC CTA GTC TCC AAA GGC GTC AAA GTC ACA      206
Met Val Gln Phe Ala Lys Arg Leu Val Ser Lys Gly Val Lys Val Thr
      35              40              45

ATC GCC ACC ACC ACC TAC ACC GCC TCC TCC ATC TCC ACT CCC TCC GTC      254
Ile Ala Thr Thr Thr Tyr Thr Ala Ser Ser Ile Ser Thr Pro Ser Val
      50              55              60

TCC GTC GAA CCA ATC TCC GAC GGC CAC GAC TTC ATC CTC ATA GGC GTC      302
Ser Val Glu Pro Ile Ser Asp Gly His Asp Phe Ile Leu Ile Gly Val
      65              70              75

CCC GGC GTC AGC ATC GAC GCA TAC TCC GAA TCC TTC AAG CTC AAC GGC      350
Pro Gly Val Ser Ile Asp Ala Tyr Ser Glu Ser Phe Lys Leu Asn Gly
      80              85              90

TCC GAA ACC TTA ACC CGA GTA ATC TCA AAA TTC AAA TCC ACA GAT TCA      398
Ser Glu Thr Leu Thr Arg Val Ile Ser Lys Phe Lys Ser Thr Asp Ser
      95              100              105              110

CCC ATC GAT TCA TTA GTC TAC GAC TCT TTC CTC CCG TGG GGA CTC GAA      446
Pro Ile Asp Ser Leu Val Tyr Asp Ser Phe Leu Pro Trp Gly Leu Glu
      115              120              125

GTC GCG AGA TCT AAC TCC ATC TCA GCT GCT GCT TTC TTC ACC AAC AAC      494
Val Ala Arg Ser Asn Ser Ile Ser Ala Ala Ala Phe Phe Thr Asn Asn
      130              135              140

CTC ACC GTT TGC TCT GTT CTA CGC AAA TTC GTC TCC GGT GAG TTT CCT      542
Leu Thr Val Cys Ser Val Leu Arg Lys Phe Val Ser Gly Glu Phe Pro
      145              150              155

CTC CCC GCT GAT CCC GCT TCC GCG CCG TAT CTC GTC CGT GGC TTA CCG      590
Leu Pro Ala Asp Pro Ala Ser Ala Pro Tyr Leu Val Arg Gly Leu Pro
      160              165              170

GCT TTG AGC TAC GAC GAG CTT CCT TCC TTC GTC GGA CGT CAC TCG TCG      638
Ala Leu Ser Tyr Asp Glu Leu Pro Ser Phe Val Gly Arg His Ser Ser
      175              180              185              190

```

AGC CAC GCG GAA CAC GGG AGA GTT CTT CTG AAC CAG TTC CGT AAC CAC Ser His Ala Glu His Gly Arg Val Leu Leu Asn Gln Phe Arg Asn His 195 200 205	686
GAA GAT GCT GAT TGG CTG TTC GTC AAC AGC TTC GAA GGG TTA GAG ACA Glu Asp Ala Asp Trp Leu Phe Val Asn Ser Phe Glu Gly Leu Glu Thr 210 215 220	734
CAA GGT TGT GAA GTT GGA GAA TCA GAG GCG ATG AGG GCG ACG TTG ATC Gln Gly Cys Glu Val Gly Glu Ser Glu Ala Met Arg Ala Thr Leu Ile 225 230 235	782
GGA CCT ATG ATA CCA TCT GCT TAT CTC GAC GGC CGA ATC AAA GAC GAT Gly Pro Met Ile Pro Ser Ala Tyr Leu Asp Gly Arg Ile Lys Asp Asp 240 245 250	830
AAA GCG TAC GCG TCG ACC CTC ATC AAC CCG CTC TCC GAC GAC TGT ATC Lys Gly Tyr Gly Ser Ser Leu Met Lys Pro Leu Ser Glu Glu Cys Met 255 260 265 270	878
GAG TGG TTA GAC ACT AAG CTG AGC AAG TCG GTG GTT TTT GTT TCG TTT Glu Trp Leu Asp Thr Lys Leu Ser Lys Ser Val Val Phe Val Ser Phe 275 280 285	926
GGT TCC TTT GGG ATC CTC TTT GAG AAG CAA CTC GCT GAG GTG GCA AAG Gly Ser Phe Gly Ile Leu Phe Glu Lys Gln Leu Ala Glu Val Ala Lys 290 295 300	974
GCG TTA CAA GAA TCC AAC TTT AAC TTC TTG TGG GTG ATC AAA GAA GCT Ala Leu Gln Glu Ser Asn Phe Asn Phe Leu Trp Val Ile Lys Glu Ala 305 310 315	1022
CAT ATA GCG AAG TTA CCA GAA GGG TTT GTG GAA GCT ACC AAA GAC AGA His Ile Ala Lys Leu Pro Glu Gly Phe Val Glu Ala Thr Lys Asp Arg 320 325 330	1070
GCG TTG CTT GTT TCT TGG TGT AAC CAG CTT GAG GTT TTA GCT CAT GTA Ala Leu Leu Val Ser Trp Cys Asn Gln Leu Glu Val Leu Ala His Val 335 340 345 350	1118
TCG ATA GGT TGC TTT TTG ACT CAC TGC GGT TGG AAC TCG ACG TTG GAA Ser Ile Gly Cys Phe Leu Thr His Cys Gly Trp Asn Ser Thr Leu Glu 355 360 365	1166
GGA TTG AGT TTG GGA GTT CCG ATG GTT GGT GTG CCT CAG TGG AGT GAT Gly Leu Ser Leu Gly Val Pro Met Val Gly Val Pro Gln Trp Ser Asp 370 375 380	1214
CAG ATG AAT GAT GCT AAG TTT GTG GAG GAG GTT TGG AGA GTT GGG TAT Gln Met Asn Asp Ala Lys Phe Val Glu Glu Val Trp Arg Val Gly Tyr 385 390 395	1262
AGA GCG AAG GAG GAA GCT GGG GGA GGA GTT GTG AAG AGC GAT GAG GTG Arg Ala Lys Glu Glu Ala Gly Gly Gly Val Val Lys Ser Asp Glu Val 400 405 410	1310
GTG AGG TGT TTG AGA GGA GTG ATG GAA GGA GAG AGT AGT GTG GAG ATT Val Arg Cys Leu Arg Gly Val Met Glu Gly Glu Ser Ser Val Glu Ile 415 420 425 430	1358
AGA GAG AGT TCT AAG AAG TGG AAA GAT TTG GCT GTG AAG GCG ATG AGT Arg Glu Ser Ser Lys Lys Trp Lys Asp Leu Ala Val Lys Ala Met Ser 435 440 445	1406
GAA GGA GGA AGC TCT GAT CGG AGC ATT AAT GAG TTT GTG GAG AGT CTA Glu Gly Gly Ser Ser Asp Arg Ser Ile Asn Glu Phe Val Glu Ser Leu 450 455 460	1454

GGG AAG AAA CAT TGAGAGGTAA TGAGATTTGT AAATCTTGTG TGTTTGTGTG 1506  
 Gly Lys Lys His  
 465

TGTTTGCT 1513

(2) INFORMATION FOR SEQ ID NO: 29:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 466 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

Met Ala Glu Thr Thr Thr Thr Thr Ala Thr Asn Ser Lys Gly His  
 1 5 10 15  
 Val Val Val Leu Pro Tyr Pro Val Gln Gly His Leu Asn Pro Met Val  
 20 25 30  
 Gln Phe Ala Lys Arg Leu Val Ser Lys Gly Val Lys Val Thr Ile Ala  
 35 40 45  
 Thr Thr Thr Tyr Thr Ala Ser Ser Ile Ser Thr Pro Ser Val Ser Val  
 50 55 60  
 Glu Pro Ile Ser Asp Gly His Asp Phe Ile Leu Ile Gly Val Pro Gly  
 65 70 75 80  
 Val Ser Ile Asp Ala Tyr Ser Glu Ser Phe Lys Leu Asn Gly Ser Glu  
 85 90 95  
 Thr Leu Thr Arg Val Ile Ser Lys Phe Lys Ser Thr Asp Ser Pro Ile  
 100 105 110  
 Asp Ser Leu Val Tyr Asp Ser Phe Leu Pro Trp Gly Leu Glu Val Ala  
 115 120 125  
 Arg Ser Asn Ser Ile Ser Ala Ala Ala Phe Phe Thr Asn Asn Leu Thr  
 130 135 140  
 Val Cys Ser Val Leu Arg Lys Phe Val Ser Gly Glu Phe Pro Leu Pro  
 145 150 155 160  
 Ala Asp Pro Ala Ser Ala Pro Tyr Leu Val Arg Gly Leu Pro Ala Leu  
 165 170 175  
 Ser Tyr Asp Glu Leu Pro Ser Phe Val Gly Arg His Ser Ser Ser His  
 180 185 190  
 Ala Glu His Gly Arg Val Leu Leu Asn Gln Phe Arg Asn His Glu Asp  
 195 200 205  
 Ala Asp Trp Leu Phe Val Asn Ser Phe Glu Gly Leu Glu Thr Gln Gly  
 210 215 220  
 Cys Glu Val Gly Glu Ser Glu Ala Met Arg Ala Thr Leu Ile Gly Pro  
 225 230 235 240  
 Met Ile Pro Ser Ala Tyr Leu Asp Gly Arg Ile Lys Asp Asp Lys Gly  
 245 250 255  
 Tyr Gly Ser Ser Leu Met Lys Pro Leu Ser Glu Glu Cys Met Glu Trp  
 260 265 270

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1459 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CCACGCGTCC	GAACAAAAACA	GAAACCAAGT	TCTGCAATGG	CGGAAACAAC	AACAACAACA	60
ACAGCGACCA	ACTCCAAAGG	CCACGTCGTG	GTCTTACCTT	ACCCAGTCCA	AGGCCACCTC	120
AACCCAATGG	TCCAGTTCGC	TAAACGCCTA	GTCTCCAAAG	GCGTCAAAGT	CACAATCGCC	180
ACCACCACCT	ACACCGCCTC	CTCCATCTCC	ACTCCCTCCG	TCTCCGTCGA	ACCAATCTCA	240
AAATTCAAAT	CCACAGATTC	ACCCATCGAT	TCATTAGTCT	ACGACTCTTT	CCTCCCGTGG	300
GGACTCGAAG	TCGCGAGATC	TAACTCCATC	TCAGCTGCTG	CTTTCTTCAC	CAACAACCTC	360
ACCGTTTGCT	CTGTTCTACG	CAAATTCGTC	TCCGGTGAGT	TTCCTCTCCC	CGCTGATCCC	420
GCTTCCGCGC	CGTATCTCGT	CCGTGGCTTA	CCGGCTTTGA	GCTACGACGA	GCTTCCTTCC	480

TTCGTCGGAC GTCACTCGTC GAGCCACGCG GAGCACGGGA GAGTTCTTCT GAACCAGTTC	540
CGTAACCACG AAGATGCTGA TTGGCTGTTC GTCAACGGCT TCGAAGGGTT AGAGACACAA	600
GGTTGTGAAG TTGGAGAATC AGAGGCGATG AAGGCGACGT TGATCGGACC TATGATACCA	660
TCTGCTTATC TCGACGGCCG AATCAAAGAC GATAAAGGCT ACGGCTCGAG CCTGATGAAG	720
CCGCTCTCGG AGGAGTGTAT GGAGTGGTTA GACACTAAGC TGAGCAAGTC GGTGGTTTTT	780
GTTTCGTTTG GTTCCTTTGG GATCCTCTTT GAGAAGCAAC TCGCTGAGGT GGCAAAGGCG	840
TTACAAGAAT CCAACTTTAA CTTCTTGTGG GTGATCAAAG AAGCTCATAT AGCGAAGTTA	900
CCAGAAGGGT TTGTGGAAGC TACCAAAGAC AGAGCGTTGC TTGTTTCTTG GTGTAACCAG	960
CTTGAGCTTT TAGCTCATCA ATCGATACCT TCCTTTTTCA CTCACCTCCC TTGGAACTCG	1020
ACGTTGGAAG GATTGAGTTT GGGAGTTCCG ATGTTTGGTG TGCCTCAGTG GAGTGATCAG	1080
ATGAATGATG CTAAGTTTGT GGAGGAGGTT TGGAGAGTTG GGTATAGAGC GAAGGAGGAA	1140
GCTGGGGGAG GAGTTGTGAA GAGCGATGAG GTGGTGAGGT GTTTGAGAGG AGTGATGGAA	1200
GGAGAGAGTA GTGTGGAGAT TAGAGAGAGT TCTAAGAAGT GGAAAGATTT GGCTGTGAAG	1260
GCGATGAGTG AAGGAGGAAG CTCTGATCGG AGCATTAAATG AGTTTGTGGA GAGTCTAGGG	1320
AAGAAACATT GAGAGGTAAT GAGATTTGTA AATCTTGTGT GTTTGTTGTT GTTGCTCAAG	1380
AGCATTGTAC GGAGCGGATT GTCATTCAGT AATATGAATA AACCAATTGT GATATTTTTT	1440
TCCTAAAAAA AAAAAAAA	1459

## (2) INFORMATION FOR SEQ ID NO: 31:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1588 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

CCACGCGTCC GTTCTCAGTC TTCTTCAACT GAAAACAAAC AGAAACTAAG GCTTCAAAGT	60
CACAATGGCG GAAACAACAA CAACAACAAC AGCAAAGACC AGCTCCAAAG GCCACGTCTT	120
GGTCTTACCT TACCCAGTCC AAGGCCACCT CAACCCAATG GTCCAGTTCC CTAAACGCCT	180
AGTCTCCAAA GGCCTCAAAG TCACAATCGC CACCACCACC TACACCGCCT CCTCCATCTC	240
CACCCCTCC GTCTCCGTCG AACCAATCTC CGACGGCCAC GACTTCATCC CCATAGGCGT	300
CCCCGGCGTC AGCATCGACG CCTACTCCGA ATCCTTCAAG CTCCACGGCT CCCAAACCTT	360
AACCCGCGTA ATCTCCAAAT TCAAATCCAC AGATTCCCCC ATCGATTCTT TAGTCTACGA	420
CTCTTTCCTC CCGTGGGGAC TCGAAGTCGC GAGATCCAAC TCCCTCTCAG CTGCCGCTTT	480
CTTCAACCAAC AACCTCACCG TTTGCTCTGT CCTTCGCAA TTCGCCTCCG GTGAGTTTCC	540
TCTCCCCGCT GATCCCCGTT CCGCGCCGTA TCTCGTCCGT GGCTTGCCGG TTTGAGCTA	600

61

CGACGAGCTT CCTTCCTTCG TGGGCCGTCA CTCGTCGAGC CACGCCGAGC ACGGGAGAGT	660
TCTTCTGAAC CAGTTCATTA ACCATGAAGA TGCTGATTGG CTGTTCTGCA ACGGCTTCGA	720
AGGGTTAGAG ACACAAGGTT GTGAAGTTGG AGAATCAGAG GCGATGAAGG CGACGTTGAT	780
CGGACCTATG ATCCCATCTG CTTATCTTGA CGCCCGAATC AAAGACGATA AAGGCTACGG	840
CTCGAGTCTG ATGAACCCGC TCTCGGAGGA GTGTATGGAG TGGTTAGACA CTAAGCTGAG	900
TAAGTCGGTG GTTTTTGTTT CGTTTGTTTC CTTTGGGATC CTCTTTGAGA AGCAACTAGC	960
TGAGGTAGCA ACGGCGTTAC AAGAATCCAA CTTTAACTTC TTGTGGGTGA TTAAAGAAGC	1020
TCATATAGCG AAGTTACCAG AAGGGTTTGT GGAAGCTACC AAAGACAGAG CGTTGCTTGT	1080
TTCTTCCTCT ACCAGCTTC AATTTTCTCT CATCATCTA TATTTGTTTT TTGACTCACT	1140
GCGGTTGGAA CTCGACGCTG GAAGGTTGAG TTTGGGAGTT CCGATGGTGG GTGTGCCGCA	1200
GTGGAGTGAT CAGATGAATG ATGCTAAGTT TGTGGAGGAG GTTTGGAGAG TTGGGTATAG	1260
AGCGAAAGAG GAAGCTGGGG GAGGAGTTGT GAAGAGCGAT GAGGTGGTGA GGTGTTTGAA	1320
AGGAGTGATG GAAGGAGAGA GTAGTGTTGA GATTAGAGAA AGTTCTAAGA AATGGAAAGA	1380
TTTGGCTGTG AAGGCCATGA GTGAAGGAGG AAGCTCTGAT CGGAGCATT AAGAGTTTGT	1440
TGAGAGTTTA GGGAAGAAAC ATTGAGAGGT AACGAGATTT GTAAATCTTG TGTGTGTTAT	1500
TGTTGTTGCT CAAGAGCATT GTACGGAGAT GATTGTCATT CAGTAATATG AATAAACCAA	1560
TTGTGATAAA AAAAAAAAAA AAAAAAAAAA	1588

## (2) INFORMATION FOR SEQ ID NO: 32:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "degenerate primer"

## (iii) HYPOTHETICAL: YES

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /mod\_base= i  
/note= "n at position 3 is inosine"

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 18
- (D) OTHER INFORMATION: /mod\_base= i  
/note= "n at position 18 is inosine"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

CCNATGATHG AYAGYGCNTA Y

21

## (2) INFORMATION FOR SEQ ID NO: 33:



- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 24 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
  - (A) DESCRIPTION: /desc = "degenerate primer"
- (ix) FEATURE:
  - (A) NAME/KEY: modified\_base
  - (B) LOCATION:4
  - (D) OTHER INFORMATION:/mod\_base= i  
/note= "n at position 4 is inosine"
- (ix) FEATURE:
  - (A) NAME/KEY: modified\_base
  - (B) LOCATION:7
  - (D) OTHER INFORMATION:/mod\_base= i  
/note= "n at position 7 is inosine"
- (ix) FEATURE:
  - (A) NAME/KEY: modified\_base
  - (B) LOCATION:13
  - (D) OTHER INFORMATION:/mod\_base= i  
/note= "n at position 13 is inosine"
- (ix) FEATURE:
  - (A) NAME/KEY: modified\_base
  - (B) LOCATION:19
  - (D) OTHER INFORMATION:/mod\_base= i  
/note= "n at position 19 is inosine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

MAANGTNGCY TCNACRAANC CYTC

24

(2) INFORMATION FOR SEQ ID NO: 34:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1969 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Brassica napus
- (ix) FEATURE:
  - (A) NAME/KEY: intron
  - (B) LOCATION:882..1054
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION:join(213..881, 1055..1783)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

CCAAATGTAT GGAAGTAGCA TTAGTTAGGT GACACCAACG CGAAACGTGA ACAATGCGGT  
CGGACGTATA CAATTATCCC CAACCACTCC ATTTTCTCC GAACACATCA GACTTATTAT

60

120

TTTTCTCCGA ACACATCAGA CTTATTATTT TTCTTCTTCC TCCTCTTCTG TGTCTTTGTC	180
AACTGCAAAC AAAACAGAAA' CCAAGTTCTG CA ATG GCG GAA ACA ACA ACA ACA Met Ala Glu Thr Thr Thr Thr 1 5	233
ACA ACA GCG ACC AAC TCC AAA GGC CAC GTC GTG GTC TTA CCT TAC CCA Thr Thr Ala Thr Asn Ser Lys Gly His Val Val Val Leu Pro Tyr Pro 10 15 20	281
GTC CAA GGC CAC CTC AAC CCA ATG GTC CAG TTC GCT AAA CGC CTA GTC Val Gln Gly His Leu Asn Pro Met Val Gln Phe Ala Lys Arg Leu Val 25 30 35	329
TCC AAA GGC GTC AAA GTC ACA ATC GCC ACC ACC ACC TAC ACC GCC TCC Ser Lys Gly Val Lys Val Thr Ile Ala Thr Thr Thr Tyr Thr Ala Ser 40 45 50 55	377
TCC ATC TCC ACT CCC TCC GTC TCC GTC GAA CCA ATC TCC GAC GGC CAC Ser Ile Ser Thr Pro Ser Val Ser Val Glu Pro Ile Ser Asp Gly His 60 65 70	425
GAC TTC ATC CCC ATA GGC GTC CCC GGC GTC AGC ATC GAC GCA TAC TCC Asp Phe Ile Pro Ile Gly Val Pro Gly Val Ser Ile Asp Ala Tyr Ser 75 80 85	473
GAA TCC TTC AAG CTC AAC GGC TCC GAA ACC TTA ACC CGA GTA ATC TCA Glu Ser Phe Lys Leu Asn Gly Ser Glu Thr Leu Thr Arg Val Ile Ser 90 95 100	521
AAA TTC AAA TCC ACA GAT TCA CCC ATC GAT TCA TTA GTC TAC GAC TCT Lys Phe Lys Ser Thr Asp Ser Pro Ile Asp Ser Leu Val Tyr Asp Ser 105 110 115	569
TTC CTC CCG TGG GGA CTC GAA GTC GCG AGA TCT AAC TCC ATC TCA GCT Phe Leu Pro Trp Gly Leu Glu Val Ala Arg Ser Asn Ser Ile Ser Ala 120 125 130 135	617
GCT GCT TTC TTC ACC AAC AAC CTC ACC GTT TGC TCT GTT CTA CGC AAA Ala Ala Phe Phe Thr Asn Asn Leu Thr Val Cys Ser Val Leu Arg Lys 140 145 150	665
TTC GCC TCC GGT GAG TTT CCT CTC CCC GCT GAT CCC GCT TCC GCG CCG Phe Ala Ser Gly Glu Phe Pro Leu Pro Ala Asp Pro Ala Ser Ala Pro 155 160 165	713
TAT CTC GTC CGT GGC TTG CCG GCT TTG AGC TAC GAC GAG CTT CCT TCC Tyr Leu Val Arg Gly Leu Pro Ala Leu Ser Tyr Asp Glu Leu Pro Ser 170 175 180	761
TTC GTG GGA CGT CAC TCG TCG AGC CAC GCG GAA CAC GGG AGA GTT CTT Phe Val Gly Arg His Ser Ser Ser His Ala Glu His Gly Arg Val Leu 185 190 195	809
CTG AAC CAG TTC CGT AAC CAC GAA GAT GCT GAT TGG CTG TTC GTC AAC Leu Asn Gln Phe Arg Asn His Glu Asp Ala Asp Trp Leu Phe Val Asn 200 205 210 215	857
GGT TTC GAA GGG TTA GAG ACA CAA GTAAGAGAAG TGTTTTAATC AAACACTCAG Gly Phe Glu Gly Leu Glu Thr Gln 220	911
TTAATAATCT ATTTTCTCAG ATTATTATTA TTATAAAAAG TAATGTATAA TTTATCTTTT	971
ATGTTTCGTTG ATGTTTTAAT TAAAATTATT TATAAACTAA TACATTCATG TTCGCTGATG	1031

TTTTGGATTG TGTTTGGTTT CAG GGT TGT GAA GTT GGA GAA TCA GAG GCG	1081
Gly Cys Glu Val Gly Glu Ser Glu Ala	
225 230	
ATG AAG GCG ACG TTG ATC GGA CCT ATG ATA CCA TCT GCT TAT CTC GAC	1129
Met Lys Ala Thr Leu Ile Gly Pro Met Ile Pro Ser Ala Tyr Leu Asp	
235 240 245	
GGC CGA ATC AAA GAC GAT AAA GGC TAC GGC TCG AGC CTG ATG AAG CCG	1177
Gly Arg Ile Lys Asp Asp Lys Gly Tyr Gly Ser Ser Leu Met Lys Pro	
250 255 260	
CTC TCG GAG GAG TGT ATG GAG TGG TTA GAC ACT AAG CTG AGC AAG TCA	1225
Leu Ser Glu Glu Cys Met Glu Trp Leu Asp Thr Lys Leu Ser Lys Ser	
265 270 275 280	
CTC CTT TTT CTT TCC TTT CTT TCC TTT GGG ATC CTC TTT GAG AAG CAA	1273
Val Val Phe Val Ser Phe Gly Ser Phe Gly Ile Leu Phe Glu Lys Gln	
285 290 295	
CTC GCT GAG GTG GCA AAG GCG TTA CAA GAA TCC AAC TTT AAC TTC TTG	1321
Leu Ala Glu Val Ala Lys Ala Leu Gln Glu Ser Asn Phe Asn Phe Leu	
300 305 310	
TGG GTG ATC AAA GAA GCT CAT ATA GCG AAG TTA CCA GAA GGG TTT GTG	1369
Trp Val Ile Lys Glu Ala His Ile Ala Lys Leu Pro Glu Gly Phe Val	
315 320 325	
GAA GCT ACC AAA GAC AGA GCG TTG CTT GTT TCT TGG TGT AAC CAG CTT	1417
Glu Ala Thr Lys Asp Arg Ala Leu Leu Val Ser Trp Cys Asn Gln Leu	
330 335 340	
GAG GTT TTA GCT CAT GAA TCG ATA GGT TGC TTT TTG ACT CAC TGC GGT	1465
Glu Val Leu Ala His Glu Ser Ile Gly Cys Phe Leu Thr His Cys Gly	
345 350 355 360	
TGG AAC TCG ACG TTG GAA GGA TTG AGT TTG GGA GTT CCG ATG GTT GGT	1513
Trp Asn Ser Thr Leu Glu Gly Leu Ser Leu Gly Val Pro Met Val Gly	
365 370 375	
GTG CCT CAG TGG AGT GAT CAG ATG AAT GAT GCT AAG TTT GTG GAG GAG	1561
Val Pro Gln Trp Ser Asp Gln Met Asn Asp Ala Lys Phe Val Glu Glu	
380 385 390	
GTT TGG AGA GTT GGG TAT AGG GCG AAG GAG GAA GCT GGG GGA GGA GTT	1609
Val Trp Arg Val Gly Tyr Arg Ala Lys Glu Glu Ala Gly Gly Gly Val	
395 400 405	
GTG AAG AGC GAT GAG GTG GTG AGG TGT TTG AGA GGA GTG ATG GAA GGA	1657
Val Lys Ser Asp Glu Val Val Arg Cys Leu Arg Gly Val Met Glu Gly	
410 415 420	
GAG AGT AGT GTG GAG ATT AGA GAG AGT TCT AAG AAG TGG AAA GAT TTG	1705
Glu Ser Ser Val Glu Ile Arg Glu Ser Ser Lys Lys Trp Lys Asp Leu	
425 430 435 440	
GCT GTG AAG GCG ATG AGT GAA GGA GGA AGC TCT GAT CGG AGC ATT AAT	1753
Ala Val Lys Ala Met Ser Glu Gly Gly Ser Ser Asp Arg Ser Ile Asn	
445 450 455	
GAG TTT GTG GAG AGT TTA GGG AAG AAA CAT TGAGAGGTAA TGAGATTTGT	1803
Glu Phe Val Glu Ser Leu Gly Lys Lys His	
460 465	
AAATCTTG TGTTTGTGTG TGTTGCTCAA GAGCATTGTA CGGAGATGAT TGTCATTCAG	1863
TAATATGAAT AAACCAATTG TGATATTTTT TTCCTAGTTC TACTGACACA CGATATGTGA	1923

AAGAATCTGC TTGTTTAAGT ACTTAGACAT GTGTATAGTT CTGCAG

1969

## (2) INFORMATION FOR SEQ ID NO: 35:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 466 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

```

Met Ala Glu Thr Thr Thr Thr Thr Thr Ala Thr Asn Ser Lys Gly His
 1           5           10           15
Val Val Val Leu Pro Tyr Pro Val Gln Gly His Leu Asn Pro Met Val
          20           25           30
Gln Phe Ala Lys Arg Leu Val Ser Lys Gly Val Lys Val Thr Ile Ala
          35           40           45
Thr Thr Thr Tyr Thr Ala Ser Ser Ile Ser Thr Pro Ser Val Ser Val
          50           55           60
Glu Pro Ile Ser Asp Gly His Asp Phe Ile Pro Ile Gly Val Pro Gly
          65           70           75           80
Val Ser Ile Asp Ala Tyr Ser Glu Ser Phe Lys Leu Asn Gly Ser Glu
          85           90           95
Thr Leu Thr Arg Val Ile Ser Lys Phe Lys Ser Thr Asp Ser Pro Ile
          100          105          110
Asp Ser Leu Val Tyr Asp Ser Phe Leu Pro Trp Gly Leu Glu Val Ala
          115          120          125
Arg Ser Asn Ser Ile Ser Ala Ala Ala Phe Phe Thr Asn Asn Leu Thr
          130          135          140
Val Cys Ser Val Leu Arg Lys Phe Ala Ser Gly Glu Phe Pro Leu Pro
          145          150          155          160
Ala Asp Pro Ala Ser Ala Pro Tyr Leu Val Arg Gly Leu Pro Ala Leu
          165          170          175
Ser Tyr Asp Glu Leu Pro Ser Phe Val Gly Arg His Ser Ser Ser His
          180          185          190
Ala Glu His Gly Arg Val Leu Leu Asn Gln Phe Arg Asn His Glu Asp
          195          200          205
Ala Asp Trp Leu Phe Val Asn Gly Phe Glu Gly Leu Glu Thr Gln Gly
          210          215          220
Cys Glu Val Gly Glu Ser Glu Ala Met Lys Ala Thr Leu Ile Gly Pro
          225          230          235          240
Met Ile Pro Ser Ala Tyr Leu Asp Gly Arg Ile Lys Asp Asp Lys Gly
          245          250          255
Tyr Gly Ser Ser Leu Met Lys Pro Leu Ser Glu Glu Cys Met Glu Trp
          260          265          270
Leu Asp Thr Lys Leu Ser Lys Ser Val Val Phe Val Ser Phe Gly Ser
          275          280          285

```

Phe Gly Ile Leu Phe Glu Lys Gln Leu Ala Glu Val Ala Lys Ala Leu  
 290 295 300  
 Gln Glu Ser Asn Phe Asn Phe Leu Trp Val Ile Lys Glu Ala His Ile  
 305 310 315 320  
 Ala Lys Leu Pro Glu Gly Phe Val Glu Ala Thr Lys Asp Arg Ala Leu  
 325 330 335  
 Leu Val Ser Trp Cys Asn Gln Leu Glu Val Leu Ala His Glu Ser Ile  
 340 345 350  
 Gly Cys Phe Leu Thr His Cys Gly Trp Asn Ser Thr Leu Glu Gly Leu  
 355 360 365  
 Ser Leu Gly Val Pro Met Val Gly Val Pro Gln Trp Ser Asp Gln Met  
 370 375 380  
 Asn Asp Ala Lys Phe Val Glu Glu Val Trp Arg Val Gly Tyr Arg Ala  
 385 390 395 400  
 Lys Glu Glu Ala Gly Gly Gly Val Val Lys Ser Asp Glu Val Val Arg  
 405 410 415  
 Cys Leu Arg Gly Val Met Glu Gly Glu Ser Ser Val Glu Ile Arg Glu  
 420 425 430  
 Ser Ser Lys Lys Trp Lys Asp Leu Ala Val Lys Ala Met Ser Glu Gly  
 435 440 445  
 Gly Ser Ser Asp Arg Ser Ile Asn Glu Phe Val Glu Ser Leu Gly Lys  
 450 455 460  
 Lys His  
 465

**CLAIMS**

1. A plant, transformed to contain a chimeric gene comprising:
  - a) a plant-expressible promoter,
  - b) a transcribed region operably linked to said promoter, comprising a DNA sequence encoding an RNA or protein, wherein said RNA or protein interfere with the normal expression of the UDP-glucose:thiohydroximate S-glucosyltransferase gene (s-gt gene) in cells of said plant, and
  - c) a 3' transcription termination and polyadenylation region active in said plant.
2. The plant of claim 1, wherein said transcribed region comprises a DNA sequence selected from the following:
  - 1) a DNA encoding an antisense RNA which is at least 80 % complementary, preferably at least 90 % complementary to the sense RNA encoded by an s-gt gene in a plant, or which has at least 80 %, preferably at least 90 %, sequence similarity to parts of said sense RNA of at least 100 nucleotides, preferably of at least 500 nucleotides;
  - 2) a DNA encoding an antisense RNA which is at least 80 % complementary, preferably at least 90 % complementary, to at least 100 nucleotides, preferably at least 500 nucleotides, of the RNA encoded by the s-gt gene, the cDNA of which is comprised in plasmid pGL9, deposited in E. coli WK6 at the BCCM-LMBP under deposit number 3344;
  - 3) a DNA encoding an antisense RNA which is at least 80 % complementary, preferably at least 90 % complementary, to at least 100 nucleotides, preferably at least 500 nucleotides, of the RNA encoded by the s-gt gene, the cDNA of which has the sequence of SEQ ID No. 28, or a DNA having substantial sequence similarity thereto; and
  - 4) a DNA encoding an antisense RNA which is at least 80 % complementary, preferably at least 90 % complementary, to at least 100 nucleotides, preferably at least 500 nucleotides, of the RNA encoded by the

s-gt gen with the sequence of SEQ ID No. 34, or a DNA having substantial sequence similarity thereto.

3. The plant of claim 1, wherein said transcribed region comprises a DNA sequence encoding a sense RNA which has at least 80 % sequence similarity, preferably at least 90 % sequence similarity, more preferably at least 95 % sequence similarity, to a region of at least 100 nucleotides, preferably a region of at least 500 nucleotides, of the following RNAs:

- 1) an RNA encoded by the s-gt gene, the cDNA of which is comprised in plasmid pGL9, deposited in E. coli WK6 at the BCCM-LMBP under deposit number 3344, or
- 2) an mRNA, the cDNA of which is shown in SEQ ID No. 28; and
- 3) an RNA encoded by the DNA of SEQ ID No. 34.

4. The plant of claim 1 wherein said transcribed region comprises a DNA sequence encoding a ribozyme with a targetting region which is at least 90%, preferably at least 95 %, complementary to part of the following RNAs:

- 1) an RNA encoded by the s-gt gene characterized by the cDNA comprised in plasmid pGL9, deposited in E. coli WK6 at the BCCM-LMBP under deposit number 3344;
- 2) an mRNA, the cDNA of which is shown in SEQ ID No. 28; and
- 3) an RNA encoded by the DNA of SEQ ID No. 34.

5. The plant of any one of claims 1 to 4, wherein said plant-expressible promoter is chosen from amongst the following group: a constitutive plant-expressible promoter, a 35S promoter, the promoter of the s-gt gene encoding an mRNA, the cDNA of which is comprised in E. coli WK6 deposited at the BCCM-LMBP under deposit number 3344, the promoter of the gene encoding an RNA, the cDNA of which has the sequence of SEQ ID No. 28, a pod tissue-specific promoter, a pod wall-specific promoter.

6. A DNA comprising a region encoding a protein with UDP-glucose:thiohydroximate S-glucosyltransferase activity, selected from the following:

- 1) a DNA encoding an mRNA, the cDNA of which is contained in plasmid pGL9, deposited in E. coli WK6 at the BCCM-LMBP under accession number 3344;
- 2) a DNA encoding an mRNA, the cDNA of which has the sequence of SEQ ID No. 28;
- 3) a DNA having substantial sequence homology to the DNA of SEQ ID No. 28; and
- 4) a DNA with the sequence of SEQ ID No. 34.

7. A DNA sequence encoding an antisense RNA selected from the following:

- 1) an antisense RNA which is complementary, preferably at least 90 % complementary, more preferably at least 95 % complementary, to a region of at least 100 nucleotides, preferably a region of at least 500 nucleotides, of an mRNA, the cDNA of which is contained in plasmid pGL9, deposited in E. coli WK6 at the BCCM-LMBP under accession number 3344;
- 2) an antisense RNA, which is at least 90 % complementary, more preferably at least 95 % complementary, to a region of at least 100 nucleotides, preferably a region of at least 500 nucleotides, of an mRNA, the cDNA of which comprises the coding region of SEQ ID No. 28;
- 3) an antisense RNA encoded by the s-gt inhibitory gene contained in plasmid pTKV8a included in E. coli MC1061, deposited at the BCCM-LMBP under accession number LMBP 3343, or an RNA having substantial sequence homology thereto; and
- 4) an antisense RNA, which is at least 90 % complementary, more preferably at least 95 % complementary, to a region of at least 100 nucleotides, preferably a region of at least 500 nucleotides, of an RNA encoded by the DNA of SEQ ID No. 34.



8. A purified plant protein, characterized by having UDP-glucose:thiohydroximate S-glucosyltransferase activity and comprising:
- 1) any of the peptide fragments of SEQ ID No. 1 to 7; or
  - 2) a region having at least 90 % sequence similarity, preferably at least 95 % sequence similarity, to any of the peptide fragments of SEQ ID Nos 1, 2, 3, 4, or 7.
9. A protein selected from the group of:
- 1) the protein encoded by the DNA contained in plasmid pGL9 deposited in E. coli WK6 at the BCCM-LMBP under accession number 3344;
  - 2) the protein with the sequence of SEQ ID No. 28;
  - 3) the protein with the sequence of SEQ ID No. 29; and
  - 4) the protein with the sequence of SEQ ID No. 35.
10. A DNA sequence encoding the protein of claims 8 or 9.
11. A plant, particularly a Brassica plant, more particularly a Brassica napus plant, transformed with any of the chimeric genes of claims 1 to 4, so to contain a level of total glucosinolates, preferably alkenyl glucosinolates, which is lower than 30  $\mu$ moles per gram dry defatted seed.
12. A Brassica napus plant, transformed so as to contain less than 30  $\mu$ moles, preferably less than 15  $\mu$ moles, particularly less than 5  $\mu$ moles, of alkenyl glucosinolates per gram of dry defatted seed.
13. A process for obtaining a Brassica napus plant having a significantly reduced expression of an s-gt gene, comprising the following steps:
- a) transforming a plant cell with any of the chimeric genes of claims 1 to 4; and
  - b) regenerating a plant from said transformed cell.

14. A process for obtaining a Brassica napus plant having having a significantly reduced content of glucosinolates in its seeds, comprising the following steps:

- a) transforming a plant cell with any of the chimeric genes of claims 1 to 4; and
- b) regenerating a plant from said transformed cell.

15. A seed, obtained from the plant of any of claims 1-4, and claims 11 or 12, comprising the chimeric gene of any of claims 1 to 4.

16. A seed cake, obtained by crushing the seed of claim 15.

17. The chimeric gene of any of claims 1 to 4.

18. A hybrid Brassica plant containing less than 30  $\mu$ moles of alkenyl glucosinolates per gram of dry defatted seed matter, obtained from two parent plants, at least one of which has total glucosinolate levels of more than 30  $\mu$ moles, preferably more than 50  $\mu$ moles, per gram dry defatted seed.

19. A hybrid Brassica plant containing less than 5  $\mu$ moles per gram of total glucosinolate levels in the whole seed basis, obtained from two parent plants, at least one of which has total glucosinolate levels of more than 5  $\mu$ moles, preferably more than 50  $\mu$ moles, per gram in the defatted seed meal.

20. A hybrid Brassica plant containing no detectable glucosinolates in the seeds, obtained from two parent plants, at least one of which has glucosinolate levels above 0  $\mu$ moles, preferably above 50  $\mu$ moles in the defatted seed meal.

# Figure 1. Schematic representation of S-GT' cDNA clones

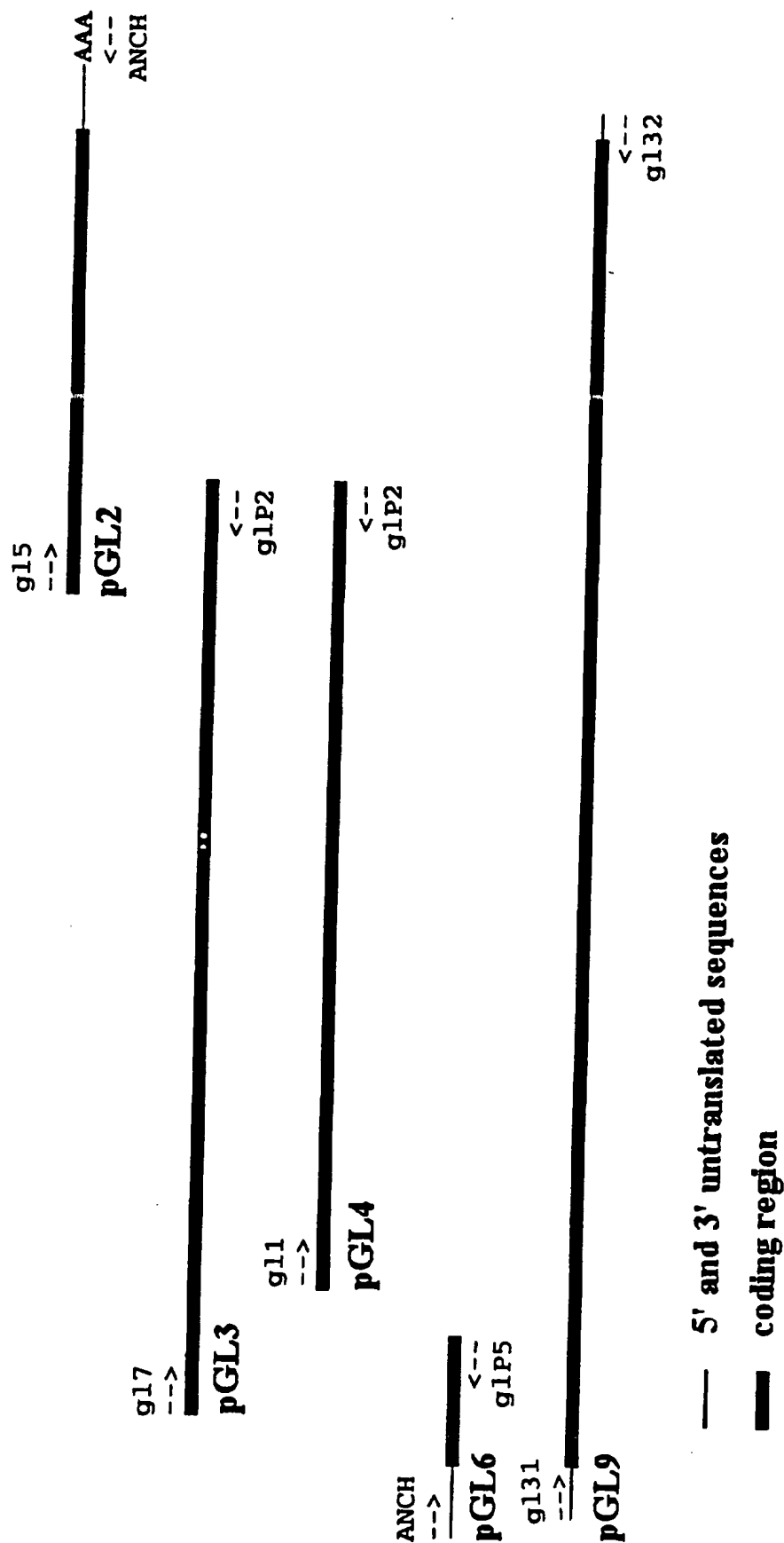


Figure 2: variants of the pGL9 clone

ACC TTA ACC CGa GTA ATC TCa AAA TTC AAA TCC ACA GAT TCa CCC ATC  
Thr Leu Thr Arg Val Ile Ser Lys Phe Lys Ser Thr Asp Ser Pro Ile  
100 105 110

3/5

T(3)

GAT TCa TTA GTC TAC GAC TCT TTC CTC CCG TGG GGA CTC GAA GTC GCG  
 Asp Ser Leu Val Tyr Asp Ser Phe Leu Pro Trp Gly Leu Glu Val Ala  
 115 120 125

C(3)

C(3)

C(3)

AGA Tct AAC TCC aTC TCA GCT Gct GCT TTC TTC ACC AAC AAC CTC ACC  
 Arg Ser Asn Ser Ile Ser Ala Ala Ala Phe Phe Thr Asn Asn Leu Thr  
 130 135 140

T(3)

C(3)

C(3)

GTT TGC TCT GTt CTa CGC AAA TTC GtC TCC GGT GAG TTT CCT CTC CCC  
 Val Cys Ser Val Leu Arg Lys Phe Val Ser Gly Glu Phe Pro Leu Pro  
 145 150 155 160

T(3)

G(3)

GCT GAT CCC GCT TCC GCG CcG TAT CTC GTC CGT GGC TTa CCG GCT TTG  
 Ala Asp Pro Ala Ser Ala Pro Tyr Leu Val Arg Gly Leu Pro Ala Leu  
 165 170 175

C(3)

G(3)

AGC TAC GAC GAG CTT CCT TCC TTC GTc GGa CGT CAC TCG TCG AGC CAC  
 Ser Tyr Asp Glu Leu Pro Ser Phe Val Gly Arg His Ser Ser Ser His  
 180 185 190

G(4)

C(3)

AT(3)

T(3)

GCg GAa CAC GGG AGA GTT CTT CTG AAC CAG TTC cGT AAC CAC GAA GAT  
 Ala Glu His Gly Arg Val Leu Leu Asn Gln Phe Arg Asn His Glu Asp  
 195 200 205

G(4)

G(3)

GCT GAT TGG CTG TTC GTC AAC aGC TTC GAA GGG TTA GAG ACA CAA GGT  
 Ala Asp Trp Leu Phe Val Asn Ser Phe Glu Gly Leu Glu Thr Gln Gly  
 210 215 220

4/5

T(3) A(4)

A(3) A(3)

TGT GAA GTT GGA GAA TCA GAG gCg ATG AgG GCG ACG TTG ATC GGA CCT  
 Cys Glu Val Gly Glu Ser Glu Ala Met Arg Ala Thr Leu Ile Gly Pro  
 225 230 235 240

C(3)

T(3) C(3)

ATG ATa CCA TCT GCT TAT CTc GAc GqC CGA ATC AAA GAC GAT AAA GGC  
 Met Ile Pro Ser Ala Tyr Leu Asp Gly Arg Ile Lys Asp Asp Lys Gly  
 245 250 255

T(3)

TAC GGC TCG AGc CTG ATG AAG CCG CTC TCG GAG GAG TGT ATG GAG TGG  
 Tyr Gly Ser Ser Leu Met Lys Pro Leu Ser Glu Glu Cys Met Glu Trp  
 260 265 270

T(3)

TTA GAC ACT AAG CTG AGc AAG TCG GTG GTT TTT GTT TCG TTT GGT TCC  
 Leu Asp Thr Lys Leu Ser Lys Ser Val Val Phe Val Ser Phe Gly Ser  
 275 280 285

A(3)

A(3)

A(3)

C(3)

TTT gGG ATC CTC TTT GAG AAG CAA CTc GCT GAG GTg GCA AaG GCG TTA  
 Phe Gly Ile Leu Phe Glu Lys Gln Leu Ala Glu Val Ala Lys Ala Leu  
 290 295 300

&lt;25

&lt;7

T(3)

G(3)

CAA GAA TCC AAC TTT AAC TTC TTG TGG GTG ATc AAA GAA GCT CaT ATA  
 Gln Glu Ser Asn Phe Asn Phe Leu Trp Val Ile Lys Glu Ala His Ile  
 305 310 315 320

GCG AAG TTA CCA GAA GGG TTT GTG GAA GCT ACC AAA GAC AGA GCG TTG  
 Ala Lys Leu Pro Glu Gly Phe Val Glu Ala Thr Lys Asp Arg Ala Leu  
 325 330 335

3&gt;

A(25)

4&gt;

G(7) T(7)

CTT GTT TCT TGG TGT AAC CAG CTT GAG GTT TTA GCT CAT GtA TCg ATA  
 Leu Val Ser Trp Cys Asn Gln Leu Glu Val Leu Ala His Val Ser Ile  
 340 345 350

5/5

T(7) C(7) G(7)  
 GGT TGc TTT TTG ACT CAC TGC GGT TGG AAC TCG ACG tTG GAA GGa TTG  
 Gly Cys Phe Leu Thr His Cys Gly Trp Asn Ser Thr Leu Glu Gly Leu  
 355 360 365

G(7) G(7)  
 AGT TTG GGA GTT CCG ATG GTt GGT GTG CCT CAG TGG AGT GAT CAG ATG  
 Ser Leu Gly Val Pro Met Val Gly Val Pro Gln Trp Ser Asp Gln Met  
 370 375 380

AAT GAT GCT AAG TTT GTG GAG GAG GTT TGG AGA GTT GGG TAT AGA GCG  
 Asn Asp Ala Lys Phe Val Glu Glu Val Trp Arg Val Gly Tyr Arg Ala  
 385 390 395 400

T(7)  
 A(7)  
AAg GAq GAA GCT GGG GGA GCA GTT GTG AAG AGC GAT GAG GTG GTG AGG  
 Lys Glu Glu Ala Gly Gly Gly Val Val Lys Ser Asp Glu Val Val Arg  
 405 410 415

A(7) A(7)  
TGT TTG AqA GGA GTG ATG GAA GGA GAG AGT AGT GTG GAG ATT AGA GAq  
 Cys Leu Arg Gly Val Met Glu Gly Glu Ser Ser Val Glu Ile Arg Glu  
 420 425 430

A(7)  
AGT TCT AAG AAg TGG AAA GAT TTG GCT GTG AAG GCG ATG AGT GAA GGA  
 Ser Ser Lys Lys Trp Lys Asp Leu Ala Val Lys Ala Met Ser Glu Gly  
 435 440 445

T(7) T(7)  
GGA AGC TCT GAT CGG AGC ATT AAT GAG TTT GTq GAG AGT cTA GGG AAG  
 Gly Ser Ser Asp Arg Ser Ile Asn Glu Phe Val Glu Ser Leu Gly Lys  
 450 455 460

25>  
 7>  
AAA CAT  
 Lys His  
 465

## INTERNATIONAL SEARCH REPORT

International Application No

PC1/EP 96/04747

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/82 C12N15/54 A01H5/00 A01H5/10

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EMBL SEQUENCE DATABASE, RELEASE 40, 13-JUL-1994. ACCESSION NUMBER L33533., XP002025912 LIM, C.O., ET AL.: "Brassica campestris pekinensis EST F0166, partial mRNA" see sequence ---	7
X	PLANT PHYSIOLOGY, vol. 105, 1994, pages 425-433, XP002001303 GROOTWASSINK, J.W.D., ET AL.: "Immunopurification and immunocharacterization of the glucosinolate biosynthetic enzyme thiohydroximate S-glucosyltransferase" cited in the application see the whole document --- -/-	8



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

## \* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

26 February 1997

Date of mailing of the international search report

11.03.97

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+31-70) 340-3016

Authorized officer

Maddox, A



## INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 96/04747

## C(Continuation) D CUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,L	WO 94 19929 A (DU PONT ;DEBONTE LORIN ROGER (US); FAN ZHEGONG (US)) 15 September 1994 * The units specified in this publication for quantifying the levels of glucosinolates are assumed as micromoles per gram, based on the accepted levels of glucosinolates in Canola, see for example W09306714 and W09010380 * see the whole document ---	12,15, 16,18,19
X	WO 93 06714 A (DU PONT) 15 April 1993  see page 4, line 4 - line 22 see page 7, line 29 - line 34 ---	12,15, 16,18,19
P,X	WO 96 00291 A (RES CORP TECHNOLOGIES INC) 4 January 1996 see example 7 ---	7
A	DNA SEQ., vol. 5, 1994, pages 41-49, XP000645061 HUGHES, J., ET AL.: "Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (Manihot esculenta Crantz) cotyledons" & EMBL SEQUENCE DATABASE, RELEASE 38, 6-FEB-1994, ACCESSION NUMBER X77464, HUGHES, J., ET AL.: "M.esculenta Crantz CGT7 mRNA for UDP-glucose glucosyltransferase" see sequence ---	8-10
A	THE PLANT JOURNAL , vol. 5, no. 1, 1994, pages 81-92, XP002026195 BRUGLIERA, F., ET AL.: "Isolation and characterisation of a cDNA clone corresponding to the Rt locus of Petunia hybrida" see page 83, right-hand column ---	7
A	WO 90 10380 A (ALLELIX INC ;WONG RAYMOND S C (CA); BEVERSDORF WALLACE D (CA); CAS) 20 September 1990 see page 1, line 23 - line 32 ---	12,15, 16,18-20

## INTERNATIONAL SEARCH REPORT

Inter:    nal Application No

PCT/EP 96/04747

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	METHODS IN PLANT BIOCHEMISTRY. CHAPTER 8, vol. 9, 1993, pages 209-237, XP000565498 POULTON, J.E., ET AL: "Glucosinolates" cited in the application see the whole document ---	1-20
A	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 91, March 1994, WASHINGTON US, pages 2166-2170, XP002001305 CHAVADEJ, S., ET AL.: "Redirection of tryptophan leads to production of low indole glucosinolate canola" cited in the application see the whole document ---	1-20
A	BULL. LIAISON - GROUPE POLYPHENOLS (1992), 16(Pt. 2), 305-8, XP000569141 VARIN, L., ET AL.: "Expression of flavonol sulfotransferase gene in canola, potato and tobacco" see the whole document ---	1-20
A	WO 87 07299 A (CALGENE INC) 3 December 1987 see page 10, line 3 - line 14 ---	1-20
A	WO 90 12084 A (DNA PLANT TECHN CORP) 18 October 1990 see page 5, line 13 - page 6, line 23 ---	1-20
A	PLANT PHYSIOLOGY, vol. 93, 1990, pages 522-524, XP002001307 LEFEBVRE, D.D.: "Expression of mammalian metallothionein suppresses glucinolate synthesis in Brassica campestris" see the whole document -----	1-20

# INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCT/EP 96/04747

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9419929 A	15-09-94	AU 6270394 A CA 2157019 A EP 0685993 A	26-09-94 15-09-94 13-12-95
WO 9306714 A	15-04-93	AU 2750692 A CA 2119859 A EP 0606359 A	03-05-93 15-04-93 20-07-94
WO 9600291 A	04-01-96	NONE	
WO 9010380 A	20-09-90	AU 642907 B	04-11-93
WO 8707299 A	03-12-87	DE 3786898 A DE 3786898 T EP 0270615 A JP 1500718 T SE 8800157 A US 5188958 A US 5463174 A	09-09-93 10-02-94 15-06-88 16-03-89 19-01-88 23-02-93 31-10-95
WO 9012084 A	18-10-90	US 5034323 A AT 123806 T AU 640644 B AU 5412390 A DE 69020151 D DE 69020151 T EP 0465572 A EP 0647715 A ES 2075897 T JP 4504800 T WO 9011682 A US 5231020 A US 5283184 A	23-07-91 15-06-95 02-09-93 05-11-90 20-07-95 28-09-95 15-01-92 12-04-95 16-10-95 27-08-92 18-10-90 27-07-93 01-02-94